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 les *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 *machines dotées d'un esprit* 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 (*computational intelligence*) est l'étude de la conception d'agents intelligents. » (Poole *et al.*, 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 [1] 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 **syllogismes** 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^{e} 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.

1.2 Fondements de l'intelligence artificielle

Cette section est consacrée à un bref historique, organisé autour d'une série de questions, des disciplines qui ont apporté des idées, des perspectives et des techniques à l'IA. Comme dans tout historique, notre propos se limite à un certain nombre de personnes, d'événements et d'idées. Bien entendu, ces problématiques ne sont pas les seules auxquelles se sont intéressées les disciplines évoquées, qui ne se sont pas développées dans la seule perspective de voir un jour éclore l'IA.

1.2.1 Philosophie

- Peut-on utiliser des règles formelles pour tirer des conclusions valides ?
- Comment l'esprit émerge-t-il à partir du cerveau physique ?
- D'où la connaissance provient-elle ?
- Comment la connaissance conduit-elle à l'action ?

Aristote (384 – 322 av. J.-C.), dont le buste apparaît sur la couverture de cet ouvrage, a été le premier à formuler un ensemble précis de lois régissant la partie rationnelle de l'esprit. Il a développé un système informel de syllogismes produisant des raisonnements valides. En principe, ce système autorisait quiconque à tirer mécaniquement des conclusions étant donné des prémisses initiales. Bien plus tard, Raymond Lulle (mort en 1315) eut l'idée qu'un artefact mécanique était capable de raisonner utilement. Thomas Hobbes (1588 – 1679) soutint que le raisonnement était analogue aux calculs sur des nombres que « nous ajoutons et soustrayons dans nos pensées silencieuses ». De son côté, l'automatisation du calcul était déjà bien avancée. Vers 1500, Léonard de Vinci (1452 – 1519) conçut sans la construire une machine à calculer dont de récentes reconstitutions ont montré qu'elle aurait pu fonctionner. La première machine à calculer connue a été fabriquée vers 1623 par le scientifique allemand Wilhelm Schickard (1592 – 1635), mais la Pascaline construite en 1642 par Blaise Pascal (1623 – 1662) est plus célèbre. Pascal a écrit que « la machine arithmétique produit des effets qui approchent plus de la pensée que tout ce que font les animaux ». Gottfried Wilhelm Leibniz (1646 – 1716) a construit une machine destinée à prendre en charge des opérations sur les concepts plutôt que sur les nombres, mais ses capacités étaient assez limitées. Leibniz a pourtant surpassé Pascal en construisant une machine capable d'additionner, de soustraire, de multiplier et d'extraire des racines carrées, alors que la Pascaline ne pouvait qu'additionner et soustraire. On se mit à spéculer sur le fait que les machines pouvaient non seulement calculer, mais également penser et agir par elles-mêmes. Dans son ouvrage *Leviathan* de 1651, Thomas Hobbes suggère l'idée d'un « animal artificiel », en argumentant ainsi : « Qu'est-ce que le cœur, sinon un ressort ; et les nerfs sinon autant de cordes ; et les articulations sinon autant de poulies. »

C'est une chose d'avancer que l'esprit fonctionne, au moins en partie, selon des règles logiques, et de construire des dispositifs physiques qui émulent certaines de ces règles ; c'en est une autre que de défendre l'idée que l'esprit lui-même *est* un tel système physique. René Descartes (1596 – 1650) a été le premier à exposer clairement la distinction entre l'esprit et la matière ainsi que les problèmes qui lui sont associés. Le problème d'une conception purement physique de l'esprit tient au fait qu'elle semble laisser peu de place au libre arbitre : si l'esprit est entièrement régi par des lois physiques, alors celui-ci a autant de liberté qu'une pierre qui « décide » de tomber en direction du centre de la Terre. Descartes était un ardent défenseur de la puissance de la raison comme outil pour comprendre le monde, une philosophie que l'on nomme aujourd'hui **rationalisme**, et qui compte Aristote et Leibniz parmi ses membres.

Mais Descartes était aussi un partisan du **dualisme**. Selon lui, une partie de l'esprit humain (ou âme) était hors de la nature et soustraite aux lois physiques. En revanche, les animaux ne possédaient pas cette qualité duale et pouvaient être considérés comme des machines. Le **matérialisme** s'oppose au dualisme, selon lequel les opérations du cerveau se conforment aux lois de la physique et *constituent* l'esprit. Le libre arbitre n'est alors plus que l'aspect sous lequel l'entité qui prend la décision perçoit les choix.

La nature physique de l'esprit qui manipule des connaissances étant établie, le problème suivant consiste à définir la source de la connaissance. Le mouvement **empiriste**, qui a commencé avec le *Novum Organum*[2] de Francis Bacon (1561 – 1626), est caractérisé par une formule de John Locke (1632 – 1704) : « Il n'y a rien dans l'entendement qui n'ait d'abord été dans les sens ». Dans son *Traité de la nature humaine* (Hume, 1739), David Hume (1711 – 1776) proposait ce qu'on appelle désormais le principe d'**induction**, selon lequel les règles générales sont élaborées à partir de la découverte d'associations répétées entre leurs éléments. S'appuyant sur les travaux de Ludwig Wittgenstein (1889 – 1951) et de Bertrand Russell (1872 – 1970), le célèbre cercle de Vienne animé par Rudolf Carnap (1891 – 1970) a développé la doctrine du **positivisme logique**. Dans cette doctrine, toute la connaissance peut être caractérisée par des théories logiques provenant, *in fine*, de **faits d'observation** qui correspondent à des perceptions sensorielles ; le positivisme logique combine donc le rationalisme et l'empirisme[3]. La **théorie de la confirmation** de Carnap et de Carl Hempel (1905 – 1997) a essayé d'analyser l'acquisition de la connaissance à partir de l'expérience. Dans *La Structure logique du monde* (1928), Carnap définit une procédure explicite de calcul permettant d'extraire des connaissances à partir d'expériences élémentaires. Il s'agit certainement de la première théorie de l'esprit décrivant celui-ci comme un processus calculatoire.

Le dernier élément de la conception philosophique de l'esprit est le lien entre la connaissance et l'action. Cette question est vitale pour l'IA, car l'intelligence requiert autant d'action que de raisonnement. En outre, ce n'est qu'en comprenant comment les actions sont justifiées que l'on peut découvrir comment construire un agent dont les actions sont justifiables (ou rationnelles). Dans *De Motu Animalium*, Aristote a défendu le fait que les actions sont justifiées par un lien logique entre des objectifs et la connaissance du résultat des actions :

> Mais comment se fait-il que la pensée soit parfois accompagnée d'une action et parfois non, parfois d'un mouvement et parfois non ? Il semble qu'il se passe la même chose lorsqu'on raisonne et qu'on produit des inférences à propos d'objets qui ne changent pas. Cependant, dans ce cas, la fin est une proposition spéculative [...] tandis que dans l'autre la conclusion produite par les deux prémisses est une action [...] J'ai besoin d'une couverture ; un manteau est une couverture ; j'ai besoin d'un manteau. Ce dont j'ai besoin, je dois le faire ; j'ai besoin d'un manteau, je dois faire un manteau. Et la conclusion « Je dois faire un manteau » est une action.

Dans l'*Éthique à Nicomaque* (Livre III, 3, 1112b), Aristote approfondit ce sujet et suggère une méthode :

> Nous délibérons non sur les fins, mais sur les moyens. En effet, ni le médecin ne délibère pour savoir s'il doit guérir, ni l'orateur pour savoir s'il doit persuader... Mais, ayant posé en

2. Le *Novum Organum* est une nouvelle version de l'*Organon* (ou instrument de la pensée) d'Aristote, lequel peut donc être considéré à la fois comme un empiriste et un rationaliste.

3. Dans ce cadre, tous les énoncés sont susceptibles d'être vérifiés ou invalidés soit par l'expérimentation, soit par l'analyse de la signification des mots. Comme ces règles rejetaient l'essentiel de la métaphysique – et telle était bien l'intention – le positivisme logique fut décrié dans certains cercles.

> principe la fin, ils examinent comment, c'est-à-dire par quels moyens, elle sera réalisée. Et s'il se révèle possible de l'obtenir par plusieurs moyens, ils examinent par lequel elle le sera le plus facilement et le mieux. Si au contraire elle ne peut être accomplie que par un seul moyen, ils examinent *comment* elle sera obtenue par ce moyen, et *ce moyen lui-même*, par quel moyen on l'obtiendra, jusqu'à ce qu'ils arrivent à la première cause, [...] et ce qu'on trouve en dernier lieu dans l'ordre de l'analyse, c'est ce qu'on fait en premier lieu dans l'ordre de réalisation. [...] [4]

L'algorithme suggéré par Aristote a été implémenté deux mille trois cents ans plus tard par Newell et Simon dans leur programme GPS. De nos jours, on le qualifierait de système de planification par régression (voir le chapitre 10).

L'analyse par objectifs est utile mais ne dit pas quoi faire lorsque plusieurs actions sont envisageables ou lorsque aucune ne permet d'atteindre complètement le but. Antoine Arnauld (1612 – 1694) a fourni dans *La Logique de Port-Royal* une description correcte d'une formule quantitative pour décider de l'action à entreprendre dans de tels cas (voir chapitre 16). Dans *Utilitarisme* (Mill, 1863), John Stuart Mill (1806 – 1873) défend le critère de la décision rationnelle dans toutes les sphères de l'activité humaine. La section suivante présente plus formellement la théorie de la décision.

1.2.2 Mathématiques

- Quelles sont les règles formelles qui permettent de tirer des conclusions valides ?
- Qu'est-ce qui peut être calculé ?
- Comment raisonne-t-on à partir d'informations incertaines ?

Si les philosophes sont à l'origine de certaines idées fondamentales pour l'IA, la transformation de cette dernière en une véritable science a exigé l'introduction d'une dose de formalisation mathématique dans trois domaines fondamentaux : la logique, le calcul et les probabilités.

On peut faire remonter l'idée de logique formelle aux philosophes de la Grèce antique, mais ses développements mathématiques n'ont vraiment commencé qu'avec les travaux de George Boole (1815 – 1864) qui a élaboré en détail la logique propositionnelle, ou booléenne (Boole, 1847). En 1879, Gottlob Frege (1848 – 1925) a étendu la logique de Boole afin d'y inclure des objets et des relations. Ce faisant, il a créé la logique du premier ordre que l'on connaît aujourd'hui [5]. Alfred Tarski (1902 – 1983) a inventé une théorie de la référence qui montre comment relier les objets d'une logique à des objets du monde réel.

L'étape suivante consistait à déterminer les limites de ce qu'on pouvait entreprendre avec la logique et le calcul. Le premier **algorithme** non trivial est attribué à Euclide pour sa méthode de calcul des plus grands diviseurs communs. Le mot *algorithme* (et l'idée d'étudier les algorithmes) remonte à al-Khawarizmi, mathématicien persan du IX^{e} siècle dont les écrits ont également introduit les chiffres arabes et l'algèbre en Europe. Boole et d'autres auteurs ont proposé des algorithmes pour la déduction logique et, dans le courant du XIX^{e} siècle, on s'est efforcé de formaliser des raisonnements mathématiques généraux sous forme de déductions logiques. En 1930, Kurt Gödel (1906 – 1978) montra qu'il existe une procédure effective pour démontrer tout énoncé vrai de la logique du premier ordre de Frege et de Russell, mais que cette logique ne peut capturer le principe d'induction mathématique nécessaire à la caractérisation des entiers naturels. En 1931, Gödel a montré que la notion de déduction

4. Traduction Gauthier & Jolif, Presses universitaires de Louvain, 1970. (*NdT*)

5. La notation proposée par Frege pour la logique du premier ordre – une combinaison astucieuse d'éléments textuels et géométriques – n'a jamais connu le succès.

possède des limites. Avec son **théorème d'incomplétude**, il a montré que dans toute théorie aussi expressive que celle de l'arithmétique de Peano (la théorie élémentaire des entiers naturels), il y a des énoncés vrais qui sont indécidables, dans le sens où ils n'ont pas de preuve dans la théorie.

Ce résultat fondamental peut aussi s'interpréter comme établissant la preuve qu'il existe des fonctions sur les entiers qui ne peuvent pas être représentées par un algorithme, autrement dit qui ne peuvent pas être calculées. C'est ce qui a incité Alan Turing (1912 – 1954) à tenter de caractériser avec exactitude les fonctions **calculables** – celles qu'on *peut* calculer avec un ordinateur. Cette notion est en fait légèrement problématique, car il n'est pas possible de définir de façon formelle la notion de calculabilité ou de procédure effective. Néanmoins, on considère généralement que la thèse de Church-Turing selon laquelle la machine de Turing (Turing, 1936) a la capacité de calculer toute fonction calculable fournit une définition suffisante. Turing a aussi montré qu'il existe des fonctions qu'aucune machine de Turing ne peut calculer. Par exemple, aucune machine ne peut dire *en général* si un programme retournera une réponse pour une entrée donnée ou s'il s'exécutera à l'infini.

Bien que la décidabilité et la calculabilité soient importantes pour comprendre la notion de calcul mécanique, celle de **praticabilité** (*tractability*) a eu un impact encore plus grand. De manière approximative, un problème est dit impraticable si le temps requis pour en résoudre des exemples croît exponentiellement avec la taille de ces exemples. La distinction entre croissance polynomiale et croissance exponentielle de la complexité a été pour la première fois mise en évidence au milieu des années 1960 (Cobham, 1964 ; Edmonds, 1965). Son importance est liée au fait que même des problèmes de petite taille ne peuvent pas être résolus en un temps raisonnable dès lors que la croissance est exponentielle. En conséquence, pour créer un comportement intelligent, il est préférable de s'attacher à subdiviser le problème concerné en sous-problèmes praticables.

Comment reconnaître un problème impraticable ? La théorie de la **NP-complétude** élaborée par Steven Cook (1971) et Richard Karp (1972) fournit une méthode. Cook et Karp ont montré l'existence de grandes classes de problèmes combinatoires NP-complets. Toute classe de problèmes réductible à la classe des problèmes NP-complets a de fortes chances d'être impraticable. (Bien qu'on n'ait jamais démontré que des problèmes NP-complets soient nécessairement impraticables, la plupart des théoriciens le pensent.) Ces résultats contredisent l'enthousiasme avec lequel la presse salua les premiers ordinateurs en les qualifiant de « super-cerveaux électroniques » qui étaient « plus rapides qu'Einstein ». En effet, en dépit de la vitesse croissante des ordinateurs, les systèmes intelligents seront caractérisés par une utilisation parcimonieuse des ressources. Dit crûment, le monde est un exemple de problème *extrêmement* grand ! Les travaux en IA ont contribué à expliquer les raisons pour lesquelles certains exemples de problèmes NP-complets sont difficiles tandis que d'autres sont faciles (Cheeseman *et al.*, 1991).

Outre la logique et la calculabilité, la troisième contribution des mathématiques à l'IA est la théorie des **probabilités**. L'Italien Jérôme Cardan (1501 – 1576) a été le premier à formuler l'idée des probabilités en décrivant celles-ci en termes de résultats possibles dans un contexte de paris. En 1654, Blaise Pascal (1623 – 1662) a montré dans une lettre à Pierre de Fermat (1601 – 1665) comment prédire le futur d'une suite de paris infinie et distribuer des gains moyens aux parieurs. Les probabilités devinrent rapidement une partie essentielle des sciences quantitatives en permettant de traiter des mesures incertaines et des théories incomplètes. James Bernoulli (1654 – 1705), Pierre-Simon Laplace (1749 – 1827) et d'autres firent avancer la

théorie et introduisirent de nouvelles méthodes statistiques. Thomas Bayes (1702 – 1761) a proposé une règle permettant d'actualiser des probabilités à partir d'observations. La règle de Bayes est sous-jacente à toutes les approches modernes du raisonnement en environnement incertain dans les systèmes d'IA.

1.2.3 Économie

- Comment prendre des décisions qui maximisent les gains ?
- Comment faire quand les autres risquent de ne pas coopérer ?
- Comment y parvenir alors que les gains sont susceptibles d'être éloignés dans le futur ?

Les sciences économiques ont pris leur essor en 1776 avec la publication de la *Recherche sur la nature et les causes de la richesse des nations* d'Adam Smith (1723 – 1790). Si les Grecs anciens notamment ont apporté des contributions à la pensée économique, Smith a été le premier à la considérer comme une science en formulant l'idée que les économies pouvaient être considérées comme composées d'agents individuels maximisant leur bien-être. De nombreuses personnes pensent que l'argent est l'objet des sciences économiques, mais les économistes diraient qu'ils étudient plutôt la manière dont les gens font des choix conduisant à la maximisation de leurs gains. Le traitement mathématique de la maximisation des gains, ou **utilité**, a été d'abord formalisé par Léon Walras (1834 – 1910). Il a été ensuite amélioré par Franck Ramsey (1931), puis par John von Neumann et Oskar Morgenstern dans *Théorie des jeux et comportement économique* (1944).

La **théorie de la décision**, qui combine la théorie des probabilités et celle de l'utilité, fournit un cadre formel complet pour les décisions (économiques ou autres) en environnement incertain, autrement dit dans les cas où des descriptions probabilistes peuvent rendre compte de l'environnement du preneur de décision. Elle convient bien aux économies de « grande dimension » dans lesquelles les agents ne tiennent pas compte des actions des autres agents considérés en tant qu'individus. Pour de « petites » économies, la situation ressemble beaucoup plus à celle d'un **jeu** : les actions d'un joueur peuvent exercer une influence considérable sur l'utilité d'un autre (positivement ou négativement). La **théorie des jeux** développée par von Neumann et Morgenstern (voir également Luce et Raiffa, 1957) contenait un résultat surprenant : dans certains jeux, un agent rationnel devrait adopter une politique qui soit aléatoire (ou au moins apparaisse comme telle). À l'inverse de la théorie de la décision, la théorie des jeux ne donne pas une méthode certaine pour sélectionner les actions.

Pour l'essentiel, les économistes n'ont pas répondu à la troisième question précédemment évoquée, à savoir comment prendre des décisions rationnelles lorsque les perspectives de gains sont éloignées et dépendent de plusieurs actions réalisées *en séquence*. Ce sujet a été étudié dans le domaine de la **recherche opérationnelle**, apparu pendant la Seconde Guerre mondiale à l'occasion des efforts entrepris en Grande-Bretagne pour optimiser les installations radars avant de trouver des applications civiles pour la prise de décisions de gestion complexes. Le travail de Richard Bellman (1957) a formalisé une classe de problèmes de décisions séquentielles – les **processus de décision de Markov** – qui sont étudiés aux chapitres 17 et 21.

Les travaux en économie et en recherche opérationnelle ont beaucoup contribué à notre notion d'agent rationnel, bien que la recherche en IA se soit développée pendant de nombreuses années en empruntant des chemins complètement distincts. Une raison tenait à l'apparente complexité de la prise de décisions rationnelles. Herbert Simon (1916 – 2001), le pionnier de la recherche en IA, a obtenu en 1978 le prix Nobel d'économie pour des tra-

vaux montrant que les modèles de prise de décision fondés sur le choix le plus **satisfaisant** (***satisficing***) – prendre des décisions « suffisamment bonnes », plutôt que sur le calcul laborieux d'un optimum – donnent une meilleure description du véritable comportement humain (Simon, 1947). On a assisté depuis les années 1990 à une résurgence de l'intérêt à l'égard des techniques de la théorie de la décision pour les systèmes d'agents (Wellman, 1995).

1.2.4 Neurosciences

– Comment le cerveau traite-t-il l'information ?

Les **neurosciences** étudient le système nerveux et en particulier le cerveau. Bien que la manière exacte dont le cerveau engendre la pensée fasse partie des grands mystères de la science, on admet depuis des millénaires que le cerveau *est*, d'une manière ou d'une autre, à la source de la pensée, pour avoir observé la diminution des capacités mentales consécutive à de violents coups sur la tête. On sait également depuis longtemps que les cerveaux humains sont un peu particuliers ; vers 335 av. J.-C., Aristote écrivait : « De tous les animaux, l'homme a le cerveau le plus important proportionnellement à sa taille [6]. » Ce n'est qu'au milieu du XVIII[e] siècle que le cerveau fut reconnu comme le siège de la conscience. Auparavant, les emplacements envisagés étaient notamment le cœur et la rate.

En 1861, l'étude par Paul Broca (1824 – 1880) de l'aphasie (un trouble du langage) chez des patients dont le cerveau avait été endommagé démontra l'existence de localisations cérébrales prenant en charge des fonctions cognitives spécifiques. Elle a en particulier montré que la production du discours articulé est localisée dans la partie de l'hémisphère gauche désormais appelée « aire de Broca » [7]. On savait à l'époque que le cerveau est composé de cellules nerveuses, ou **neurones**, mais ce n'est qu'en 1873 que Camillo Golgi (1843 – 1926) développa une technique de coloration qui permit d'observer des neurones individuels (voir figure 1.2). Cette technique a été utilisée par Santiago Ramón y Cajal (1852 – 1934) dans ses études pionnières sur les structures neuronales du cerveau [8]. Nicolas Rashevsky (1936, 1938) a été le premier à appliquer des modèles mathématiques à l'étude du système nerveux.

On connaît désormais la correspondance entre les zones du cerveau et les parties du corps humain qu'elles contrôlent ou desquelles elles reçoivent des *stimuli* sensoriels. Ces correspondances peuvent changer complètement en quelques semaines et certains animaux semblent avoir plusieurs systèmes de correspondance. En outre, on ne comprend pas parfaitement comment d'autres zones peuvent compenser les fonctions d'une zone endommagée. Enfin, il n'existe pratiquement pas de théorie quant à la conservation de la mémoire individuelle.

La mesure de l'activité d'un cerveau intact a commencé en 1929 avec l'invention par Hans Berger (1873 – 1941) de l'électroencéphalogramme (EEG). Les récentes avancées de l'imagerie par résonance magnétique fonctionnelle, ou IRMf (Ogawa *et al.*, 1990 ; Cabeza et Nyberg, 2001), donnent aux neuroscientifiques des images détaillées de l'activité du cerveau sans précédent, ce qui permet de réaliser des mesures correspondant aux processus cognitifs en cours. Celles-ci sont complétées par les progrès accomplis dans l'observation de l'activité des neurones au niveau de la cellule. On peut stimuler électriquement, chimiquement ou

6. On a découvert depuis que certains petits mammifères (les scandentiens) ont un rapport masse du cerveau à masse de l'organisme plus élevé que le nôtre.

7. Alexander Hood (1824) est souvent cité comme un précurseur possible.

8. Golgi s'est entêté à croire que les fonctions du cerveau étaient principalement prises en charge par un tissu continu dans lequel les neurones étaient enchâssés, tandis que Cajal soutenait la « doctrine neuronale ». Tous deux ont partagé le prix Nobel en 1906 mais ont prononcé des discours d'acceptation antagonistes.

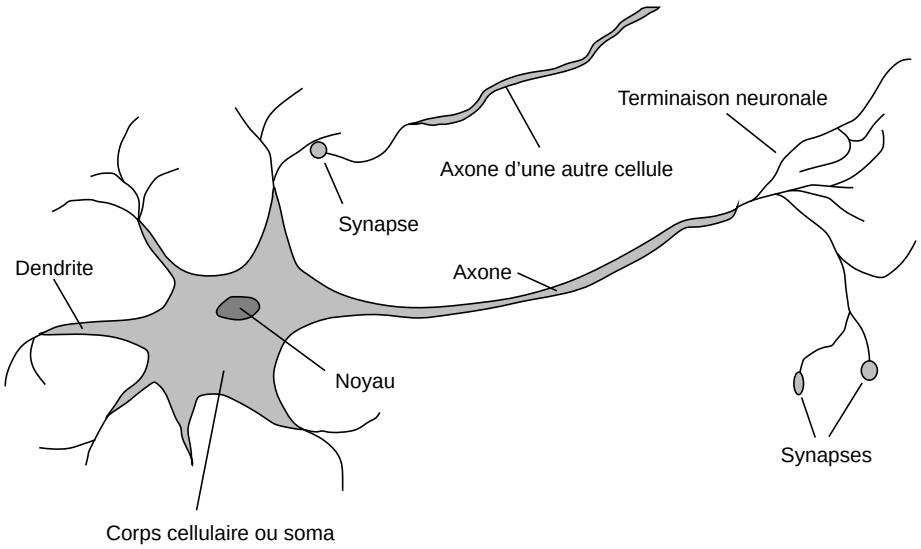

Figure 1.2 : Les constituants d'une cellule nerveuse ou neurone. Chaque neurone est composé d'un corps cellulaire, ou soma, qui contient un noyau. Du corps cellulaire sont issues plusieurs fibres, appelées *dendrites*, et une longue fibre, appelée *axone*. L'axone est beaucoup plus long que ce qu'indique le schéma. Un axone mesure généralement 1 cm (100 fois le diamètre d'un corps cellulaire), mais il peut atteindre 1 m. Un neurone est connecté avec 10 à 100 000 autres neurones par des points de contact appelés *synapses*. Les signaux se propagent d'un neurone à l'autre par une réaction électrochimique complexe ; ils contrôlent l'activité du cerveau à court terme tout en permettant des modifications à long terme dans la connectivité des neurones. On pense que ces mécanismes forment la base de l'apprentissage dans le cerveau. L'essentiel du traitement des informations a lieu dans le cortex cérébral, l'« écorce » du cerveau. L'unité organisationnelle de base semble être une colonne de tissu d'environ 0,5 mm de diamètre, contenant environ 20 000 neurones et traversant le cortex, lequel a une profondeur d'à peu près 4 mm chez l'homme.

même optiquement (Han et Boyden, 2007), des neurones individuellement, on peut ainsi cartographier des relations d'entrées-sorties à ce niveau. En dépit de ces progrès, il reste beaucoup de chemin à parcourir avant de comprendre le véritable fonctionnement des processus cognitifs.

Le plus surprenant, c'est qu'*un ensemble de cellules élémentaires puisse être à l'origine de la pensée, de l'action et de la conscience* ou, selon l'aphorisme de John Searle (1992), que des *cerveaux causent des esprits*. La seule véritable alternative est le mysticisme : l'idée qu'il y aurait une dimension mystique située au-delà de la science physique dans laquelle les esprits fonctionneraient.

Le cerveau et l'ordinateur ont des propriétés assez différentes. La figure 1.3 montre que les ordinateurs ont un temps de cycle un million de fois plus rapide que le cerveau, lequel compense par une capacité mémoire et un nombre d'interconnexions bien plus élevés que ce dont dispose un ordinateur individuel haut de gamme, et ce, même si les plus gros super-ordinateurs ont des capacités comparables à celles du cerveau. (On doit remarquer cependant, que le cerveau ne semble pas faire usage de tous ses neurones simultanément.) Les futurologues se régalent avec ces nombres, en prédisant que nous allons atteindre bientôt une **singularité**, un point à partir duquel les ordinateurs dépasseront les humains en performances (Vinge, 1993 ; Kurzweil, 2005), mais les performances brutes ne mesurent pas grand-chose.

	Superordinateur	Ordinateur personnel	Cerveau humain
Unités de traitement	10^4 CPU 10^{12} transistors	4 CPU 10^9 transistors	10^{11} neurones
Unités de stockage	10^{14} bits en RAM 10^{15} bits sur disque	10^{11} bits en RAM 10^{13} bits sur disque	10^{11} neurones 10^{14} synapses
Durée des cycles	10^{-9} secondes	10^{-9} secondes	10^{-3} secondes
Opérations/sec	10^{15}	10^{10}	10^{17}
Transactions mémoire/sec	10^{14}	10^{10}	10^{14}

Figure 1.3 : Comparaison sommaire des ressources brutes dont disposent le superordinateur IBM BLUE GENE, un ordinateur personnel typique de 2008, et le cerveau humain. Les données relatives au cerveau humain restent constantes, alors que celles du superordinateur ont été multipliées par un facteur 10 tous les cinq ans environ, lui permettant d'arriver au niveau du cerveau. L'ordinateur personnel reste loin derrière dans toutes les dimensions, sauf en ce qui concerne le temps de cycle.

Même si nous disposions d'un ordinateur de capacité virtuellement infinie, nous ne saurions toujours pas comment le programmer pour atteindre le niveau d'intelligence du cerveau.

1.2.5 Psychologie

– Comment les hommes et les animaux pensent et agissent-ils ?

On fait habituellement remonter les origines de la psychologie scientifique aux travaux du physicien allemand Hermann von Helmholtz (1821 – 1894) et de son disciple Wilhelm Wundt (1832 – 1920). Helmholtz a appliqué la méthode scientifique à l'étude de la vision humaine et son *Traité d'optique physiologique* est encore décrit comme « le traité le plus important sur la physique et la physiologie de la vision humaine » (Nalwa, 1993, page 15). En 1879, Wundt ouvrait le premier laboratoire de psychologie expérimentale à l'université de Leipzig. Il insistait sur les expériences soigneusement contrôlées au cours desquelles ses collaborateurs devaient réaliser une tâche perceptive ou associative tout en se livrant à une introspection dans leur processus de pensée. Si les instruments de contrôle employés contribuèrent à l'émergence de la psychologie en tant que science, la nature subjective des données rendait peu vraisemblable qu'un expérimentateur réfute un jour ses propres théories. À l'inverse, les biologistes qui étudient le comportement animal ne disposent pas des données fournies par l'introspection et ont développé une méthodologie objective décrite par H. S. Jennings (1906) dans son ouvrage majeur, *Behavior of the Lower Organisms*. C'est en appliquant ce point de vue aux humains que le mouvement **béhavioriste** dirigé par John Watson (1878 – 1958) a rejeté *toute* théorie faisant appel à des processus mentaux en raison de l'impossibilité d'obtenir des observations fiables au moyen de l'introspection. Les béhavioristes ont insisté sur la prise en compte des seules mesures objectives des percepts (ou *stimuli*) envoyés à un animal et des actions résultantes (ou *réponses*). Le béhaviorisme a découvert beaucoup de choses sur les rats et sur les pigeons mais a eu moins de succès pour ce qui est de la compréhension des humains.

La **psychologie cognitive**, qui voit le cerveau comme instance de traitement des informations, remonte au moins aux travaux de Williams James (1842 – 1910). Helmholtz insistait également sur l'idée que la perception nécessite une forme d'inférence logique inconsciente. Alors que le point de vue cognitif était grandement éclipsé par le béhaviorisme aux États-Unis,

la modélisation cognitiviste continuait à être développée à l'Unité de psychologie appliquée, dirigée par Frederic Bartlett (1886 – 1969) à Cambridge. Dans *The Nature of Explanation* (1943), Kenneth Craik, étudiant puis successeur de Bartlett, rétablit avec force la légitimité de termes « mentaux » tels que « croyances » et « buts », en affirmant qu'ils étaient tout aussi scientifiques que, par exemple, les termes de « pression » et de « température » employés à propos des gaz bien que ceux-ci soient constitués de molécules auxquelles ces propriétés ne s'appliquent pas. Craik spécifia les trois grandes étapes d'un agent fondé sur les connaissances (*knowledge-based*) : (1) le *stimulus* doit être converti en représentation interne, (2) la représentation est manipulée par des processus cognitifs de manière à dériver de nouvelles représentations, (3) celles-ci sont à leur tour transformées en actions. Il a clairement expliqué les raisons pour lesquelles ces étapes schématisaient bien un agent.

> Si l'organisme contient dans sa tête un « modèle réduit » de la réalité extérieure et de ses actions possibles, il est en mesure d'essayer différentes possibilités, de conclure laquelle est la meilleure, de réagir à des situations futures avant qu'elles ne surviennent, d'utiliser la connaissance des événements passés pour traiter le présent et le futur et, quelle que soit l'issue, de réagir d'une manière plus complète, plus sûre et plus compétente aux urgences auxquelles il est confronté. (Craik, 1943)

Après la mort de Craik dans un accident de vélo en 1945, ses recherches furent poursuivies par Donald Broadbent (1926 – 1993), dont le livre *Perception and Communication* (1958) était l'un des premiers travaux à modéliser les phénomènes psychologiques comme traitement de l'information. Dans le même temps, aux États-Unis, le développement de la modélisation informatique conduisait à la création du champ des **sciences cognitives**. On peut dire que ce champ a vu le jour au MIT, lors d'un séminaire organisé en septembre 1956. (Celui-ci s'est tenu deux mois après la conférence qui a vu « naître » l'IA.) Au cours de ce séminaire, George Miller a présenté *The Magic Number Seven ;* Noam Chomsky, *Three Models of Language ;* Allen Newell et Herbert Simon ont quant à eux exposé *The Logic Theory Machine*. Ces trois interventions importantes ont montré comment utiliser des modèles informatiques dans le cadre des problématiques de la psychologie de la mémoire, du langage et de la pensée logique, respectivement. Les psychologues admettent désormais généralement (bien que pas encore universellement) qu'« une théorie cognitive doit être comme un programme informatique » (Anderson, 1980), autrement dit qu'elle doit décrire en détail le mécanisme de traitement de l'information grâce auquel une fonction cognitive pourrait être implémentée.

1.2.6 Ingénierie informatique

– Comment construire un ordinateur performant ?

La réussite de l'intelligence artificielle nécessite deux composantes : l'intelligence et un artefact. L'ordinateur est devenu l'artefact privilégié. Des scientifiques de trois des pays belligérants de la Seconde Guerre mondiale ont inventé l'ordinateur moderne de manière indépendante et presque simultanée. L'équipe d'Alan Turing a construit en 1940 le premier ordinateur *opérationnel,* le calculateur électromécanique Heath Robinson [9], en vue d'une fonction unique : déchiffrer les messages des Allemands. En 1943, la même équipe développa le Colossus, ma-

9. Heath Robinson était un illustrateur célèbre pour ses descriptions d'ustensiles bizarres et compliqués destinés à des tâches routinières telles que beurrer une tartine.

chine puissante et polyvalente composée de tubes à vide [10]. Le Z-3, inventé par Konrad Zuse en Allemagne en 1941, fut le premier ordinateur opérationnel et *programmable*. Zuse a aussi inventé les nombres en virgule flottante et le premier langage de programmation évolué, le Plankalkül. John Atanasoff et son étudiant Clifford Berry ont assemblé le premier ordinateur *électronique*, l'ABC, entre 1940 et 1942 à l'université de l'État de l'Iowa. La recherche d'Atanasoff ne reçut que peu de soutien et de reconnaissance ; c'est l'ENIAC, développé dans le cadre d'un projet militaire secret à l'université de Pennsylvanie par une équipe incluant John Mauchly et John Eckert, qui s'est révélé le principal précurseur des ordinateurs modernes.

Depuis cette époque, chaque génération d'ordinateurs a été plus rapide, plus puissante et moins onéreuse. Les performances ont doublé à peu près tous les dix-huit mois jusque vers 2005, quand les problèmes de dissipation thermique ont conduit les fondeurs à multiplier le nombre de cœurs de CPU plutôt qu'à augmenter la vitesse d'horloge. Les prédictions actuelles montrent que la montée en puissance de calcul viendra du parallélisme massif, une curieuse coïncidence avec les propriétés du cerveau.

Bien entendu, il existait des machines à calculer bien avant l'apparition des ordinateurs électroniques. Nous avons déjà évoqué les premières machines automatisées (voir page 6). La première machine *programmable*, un métier à tisser mis au point en 1805 par Joseph-Marie Jacquard (1752 – 1834), utilisait des cartes perforées pour enregistrer les instructions associées au motif à réaliser. Au milieu du XIX^{e} siècle, Charles Babbage (1792 – 1871) conçut deux machines mais n'en construisit aucune. La « machine à différences » était destinée à calculer des tables mathématiques pour des projets d'ingénierie et scientifiques. Elle a finalement été construite et mise en œuvre en 1991 au musée des Sciences de Londres (Swade, 2000). Son autre projet, une « machine analytique » était beaucoup plus ambitieux : elle disposait d'une mémoire adressable, de programmes enregistrés et de sauts conditionnels ; il s'agissait du premier artefact capable d'effectuer des calculs en tout genres. Ada Lovelace, fille du poète Lord Byron et collègue de Babbage, a certainement été le premier programmeur du monde. (C'est pour lui rendre hommage que le nom du langage de programmation Ada a été choisi.) Elle a écrit des programmes pour la machine analytique inachevée et a même envisagé que celle-ci puisse jouer aux échecs ou composer de la musique.

L'IA doit beaucoup aussi à la partie logicielle de l'informatique qui a fourni les systèmes d'exploitation, les langages de programmation et les outils nécessaires à l'écriture des programmes modernes (ainsi qu'à leur documentation). Mais il s'agit là d'un domaine pour lequel la dette a été remboursée : les travaux en IA ont fait éclore de nombreuses idées reprises en informatique. Parmi celles-ci, on peut citer le temps partagé, les interpréteurs interactifs, les ordinateurs personnels dotés d'une interface graphique et d'une souris, les environnements de développement rapide, les listes chaînées, la gestion automatique de la mémoire et les concepts clés de la programmation symbolique, fonctionnelle, déclarative et orientée objet.

1.2.7 Théorie du contrôle et cybernétique

– Comment faire en sorte que des artefacts opèrent de façon autonome ?

Ktesibios d'Alexandrie (vers 250 av. J.-C.) a construit le premier dispositif autorégulé : une horloge à eau dotée d'un régulateur pour maintenir constant le débit. Cette invention a changé la

10. Dans la période d'après-guerre, Turing souhaita utiliser ces ordinateurs pour la recherche en IA, par exemple pour l'un des premiers programmes d'échecs (Turing *et al.*, 1953), mais le gouvernement britannique bloqua ses efforts.

définition des possibilités réalisables par un artefact. Auparavant, seuls des êtres vivants pouvaient modifier leur comportement en réponse à des changements dans leur environnement. On peut citer d'autres exemples de systèmes asservis autorégulés : le régulateur du moteur à vapeur de James Watt (1736 – 1819) et le thermostat créé par Cornelis Drebbel (1572 – 1633), qui est aussi l'inventeur du sous-marin. C'est au cours du XIX^{e} siècle qu'a été développée la théorie mathématique de la stabilité des systèmes asservis.

Norbert Wiener (1894 – 1964) a tenu une place centrale dans la création de ce qu'on appelle désormais la **théorie du contrôle** (ou **théorie de la commande**). Brillant mathématicien, Wiener a notamment travaillé avec Bertrand Russell avant de s'intéresser aux systèmes de contrôle biologiques et mécaniques et à leur rapport avec la cognition. Comme Craik (qui utilisait également des systèmes de contrôle comme modèles psychologiques), Wiener et ses collègues Arturo Rosenblueth et Julian Bigelow ont défié l'orthodoxie béhavioriste (Rosenblueth *et al.*, 1943). Selon eux, le comportement piloté par un but résultait d'un mécanisme régulateur essayant de minimiser l'« erreur » – la différence entre l'état courant et l'état but. À la fin des années 1940, Wiener, entouré de Warren McCulloch, Walter Pitts et John von Neumann, organisa une série de conférences séminales consacrées aux nouveaux modèles mathématiques et informatiques de la cognition. Le livre de Wiener, *Cybernetics* (1948), devint un best-seller qui fit prendre conscience au public des possibilités des machines artificiellement intelligentes. Pendant ce temps, en Grande Bretagne, W. Ross Ashby défrichait un champ similaire (Ashby, 1940). Ashby, Alan Turing, Grey Walter et d'autres ont formé le *Ratio Club* pour « ceux qui ont eu des idées analogues à celles de Wiener avant que son livre ne paraisse ». L'ouvrage d'Ashby, *Design for a Brain* (1948, 1952), développe l'idée selon laquelle on pourrait créer l'intelligence par l'utilisation de dispositifs **homéostatiques**, contenant des boucles de rétroaction appropriées qui permettraient de garantir un comportement adaptatif stable.

La théorie du contrôle moderne, et tout particulièrement la branche appelée *contrôle stochastique optimal*, a notamment pour but la conception de systèmes maximisant une **fonction objectif** au cours du temps. Cela correspond à peu près à notre vision de l'IA : la conception de systèmes au comportement optimal. Dans ce cas, pourquoi l'IA et la théorie de la commande forment-elles deux domaines distincts, malgré les relations étroites que leurs fondateurs entretenaient ? La réponse réside dans la forte connexion qui existe entre les techniques mathématiques qui leur étaient familières et les types de problèmes abordés par chaque domaine. Les outils de la théorie du contrôle, le calcul différentiel et l'algèbre matricielle sont plus spécifiquement destinés à des systèmes qui se décrivent sous forme d'ensembles fixes de variables continues, alors que l'IA a été fondée en partie pour échapper aux limites perceptibles de ces outils. La logique, sous sa forme informatique, a donné aux chercheurs en IA la possibilité d'étudier des problèmes comme le langage, la vision et la planification qui tombaient en dehors du champ d'investigation des puristes de la théorie de la commande.

1.2.8 Linguistique

– Quels sont les rapports entre le langage et la pensée ?

En 1957, B. F. Skinner publiait *Verbal Behavior*. Écrit par le plus éminent expert du domaine, cet ouvrage offre un panorama complet et détaillé de l'approche béhavioriste sur le sujet de l'apprentissage du langage. Mais, curieusement, un compte-rendu du livre est devenu aussi célèbre que le livre lui-même et a permis de faire pratiquement disparaître tout intérêt pour le béhaviorisme. Le linguiste Noam Chomsky, auteur du compte-rendu, venait de publier un

ouvrage exposant sa propre théorie, *Structures syntaxiques*. Il y faisait remarquer en quoi la théorie béhavioriste était impuissante à rendre compte de la notion de créativité langagière – elle n'expliquait pas comment un enfant pouvait comprendre et construire des phrases qu'il n'avait jamais entendues auparavant. La théorie de Chomsky, fondée sur des modèles syntaxiques remontant au linguiste indien Panini (vers 350 av. J.-C.), pouvait l'expliquer et, à la différence des théories précédentes, elle était suffisamment formelle pour que sa programmation soit envisageable.

La linguistique moderne et l'IA sont donc nées à la même époque et ont évolué ensemble ; elles se croisent dans un domaine hybride appelé **linguistique computationnelle** ou **traitement automatique du langage naturel**. Le problème de la compréhension du langage s'est révélé rapidement beaucoup plus complexe qu'il n'y paraissait en 1957. Outre la connaissance de la structure des phrases, celle-ci requiert l'appréhension du sujet et du contexte. Cela peut sembler évident aujourd'hui, mais ce ne fut pas le cas avant les années 1960. L'essentiel des premiers travaux sur la **représentation des connaissances** (l'étude de la traduction des connaissances sous une forme qui permette à l'ordinateur de raisonner) était lié au langage et nourri par des recherches en linguistique, laquelle puisait à son tour dans des décennies de travaux consacrés à l'analyse philosophique du langage.

1.3 Histoire de l'intelligence artificielle

1.3.1 Gestation de l'intelligence artificielle (1943 – 1955)

Les premiers travaux désormais généralement reconnus comme appartenant à l'IA ont été menés par Warren McCulloch et Walter Pitts (1943). Ils puisèrent à trois sources : l'état du savoir sur la physiologie de base et la fonction des neurones dans le cerveau, l'analyse formelle de la logique propositionnelle de Russell et Whitehead, et la théorie du calcul de Turing. Ils proposèrent un modèle de neurones artificiels dans lequel chaque neurone est caractérisé par un état « marche » ou « arrêt », le passage à l'état « marche » se produisant en réponse à une stimulation émise par un nombre suffisant de neurones voisins. L'état d'un neurone était conçu comme « de fait équivalent à une proposition présentant son *stimulus* approprié ». Ils montrèrent, par exemple, que toute fonction calculable peut être calculée par un réseau de neurones connectés et que tous les connecteurs logiques (et, ou, non, etc.) peuvent être implémentés par des structures simples. McCulloch et Pitts ont également suggéré que des réseaux définis de manière appropriée sont capables d'apprentissage. Donald Hebb a trouvé une règle de mise à jour simple, maintenant appelée **apprentissage hebbien**, qui permet de modifier les intensités des connexions entre les neurones et qui demeure un modèle très influent aujourd'hui (1949).

Deux étudiants de Harvard, Marvin Minsky et Dean Edmonds, ont construit le premier ordinateur à réseau de neurones en 1950. Le SNARC, tel était son nom, utilisait 3 000 tubes à vide et un mécanisme de pilote automatique récupéré sur un bombardier B-24 pour simuler un réseau de 40 neurones. Plus tard, à Princeton, Minsky étudia le calcul universel sur des réseaux de neurones. Le jury de thèse de Minsky émit des doutes quant à la nature mathématique de ce travail, mais on rapporte que von Neumann déclara : « Si ce n'est pas le cas aujourd'hui, ce le sera un jour. » Plus tard, Minsky démontra des théorèmes importants attestant les limites de la recherche sur les réseaux neuronaux.

Plusieurs exemples de travaux antérieurs pourraient être caractérisés comme relevant de l'IA, mais la vision d'Alan Turing a peut-être été la plus influente. Il donna des conférencs sur

le sujet dès 1947 à la *London Mathematical Society* et dégagea clairement une feuille de route convaincante dans son article de 1950 « Computing Machinery and Intelligence ». C'est dans ce texte qu'il a présenté le test de Turing, l'apprentissage artificiel, les algorithmes génétiques et l'apprentissage par renforcement. Il proposa l'idée du *Child Programme*, expliquant que « Plutôt que d'essayer de produire un programme qui simule l'esprit d'un adulte, pourquoi ne pas plutôt tenter de simuler celui d'un enfant ? ».

1.3.2 Naissance de l'intelligence artificielle (1956)

À Princeton se trouvait également John McCarthy, autre personnalité majeure de l'IA. Après avoir reçu son doctorat et avoir exercé deux ans comme professeur, McCarthy déménagea à Stanford, puis à l'université de Dartmouth qui allait devenir le lieu de naissance officiel de la discipline. McCarthy convainquit Minsky, Claude Shannon et Nathaniel Rochester de l'aider à rassembler les chercheurs américains spécialisés dans la théorie des automates, les réseaux neuronaux et l'étude de l'intelligence. Ils organisèrent un séminaire de deux mois à Dartmouth au cours de l'été 1956. Le projet stipulait [11] :

> Nous proposons d'entreprendre une étude de l'intelligence artificielle à dix personnes pendant deux mois durant l'été 1956 à Dartmouth College, Hanover dans le New Hampshire. L'étude reposera sur la conjecture que chaque aspect de l'apprentissage ou de toute autre facette de l'intelligence peut être décrit en principe si précisément qu'une machine puisse être construite pour le simuler. On tentera de proposer des solutions pour que les machines puissent utiliser le langage, former des abstractions et des concepts, résoudre des types de problèmes réservés aux humains pour l'instant, et se perfectionner. Nous pensons que des avancées significatives sont possibles si une équipe judicieusement sélectionnée de scientifiques travaille sur ce sujet pendant un été.

Une dizaine de personnes y assistèrent, dont Trenchard More de Princeton, Arthur Samuel d'IBM ainsi que Ray Solomonoff et Oliver Selfridge du MIT.

Deux chercheurs de Carnegie Tech [12], Allen Newell et Herbert Simon, volèrent la vedette lors de cet événement. En effet, si les autres participants avaient des idées et parfois même des programmes pour des applications particulières telles que le jeu de dames, Newell et Simon disposaient déjà d'un programme capable de raisonner, le *Logic Theorist* (LT), duquel Simon disait : « Nous avons inventé un programme informatique capable de penser de manière non numérique et, ce faisant, de résoudre le vénérable problème de la dualité du corps et de l'esprit [13]. » Peu après le séminaire, le programme était en mesure de démontrer la majorité des théorèmes du chapitre 2 des *Principia Mathematica* de Russell et Whitehead. Il paraît que Russell fut ravi lorsque Simon lui montra que le programme avait trouvé, pour un théorème, une démonstration plus courte que celle présentée dans le livre. Les éditeurs du *Journal of*

11. C'était la première utilisation des termes *intelligence artificielle* de McCarthy. Peut-être que « rationalité computationnelle » aurait été plus précis et moins menaçant mais « IA » a prévalu. Au 50e anniversaire de la conférence de Dartmouth, McCarty a expliqué qu'il avait renoncé aux termes « ordinateur » ou « computationnel » par déférence envers Norbert Wiener, qui promouvait des dispositifs cybernétiques analogiques, plutôt que les ordinateurs numériques.

12. Désormais Carnegie Mellon University (CMU).

13. Newell et Simon ont aussi inventé le langage de traitement de listes IPL pour écrire LT. Comme ils n'avaient pas de compilateur, ils convertirent les instructions en code machine à la main. Pour éviter les erreurs, ils travaillaient en parallèle, chacun demandant à l'autre les nombres binaires à mesure afin de s'assurer qu'ils étaient d'accord.

Symbolic Logic furent moins impressionnés ; ils rejetèrent un papier coécrit par Newell, Simon et le *Logic Theorist*.

Le séminaire de Dartmouth n'a pas conduit à de nouvelles avancées mais il a au moins permis aux principaux acteurs de l'IA de faire connaissance. Au cours des deux décennies suivantes, ces personnalités et leurs étudiants ou collègues du MIT, de Carnegie Mellon, de Stanford et d'IBM ont dominé la discipline.

C'est en lisant le projet de séminaire de Dartmouth (McCarthy *et al.*, 1955) qu'on comprend que l'IA devait devenir une discipline distincte. Pourquoi tous les travaux effectués en IA ne pouvaient-ils pas se rattacher à la théorie du contrôle, à la recherche opérationnelle ou à la théorie de la décision, qui poursuivent après tout des objectifs analogues à ceux de l'IA ? Ou pourquoi l'IA n'est-elle pas une branche des mathématiques ? La première réponse tient au fait que, dès le départ, l'IA a eu l'ambition de dupliquer des facultés humaines telles que la créativité, l'apprentissage et l'utilisation du langage. Aucune des autres disciplines n'envisageait ces problèmes. La seconde réponse est d'ordre méthodologique. L'IA est la seule discipline qui soit clairement une branche de l'informatique (même si la recherche opérationnelle insiste elle aussi sur les simulations informatiques) et qui s'attache à construire des machines fonctionnant de manière autonome dans des environnements en évolution.

1.3.3 L'enthousiasme des débuts : les grandes espérances (1952 – 1969)

Les premières années de l'IA furent marquées par un grand nombre de succès, que l'on peut maintenant relativiser. Compte tenu du caractère rudimentaire des ordinateurs et des outils de programmation de l'époque, et sachant qu'on considérait peu de temps auparavant que les ordinateurs ne pouvaient faire que de l'arithmétique et rien d'autre, on s'émerveillait dès qu'une machine accomplissait quelque chose d'un peu évolué. On admettait très largement dans les milieux intellectuels qu'« une machine ne pourrait jamais faire *X* » (voir au chapitre 26 la longue liste des *X* collectionnés par Turing). Naturellement, les chercheurs en IA ont répondu en faisant la démonstration d'un *X* après l'autre. John McCarthy définit cette période comme celle des premiers pas : « Hé, maman… t'as vu ? Sans les mains ! »

Le premier succès de Newell et de Simon fut suivi du *General Problem Solver*, ou GPS. À la différence de *Logic Theorist*, ce programme a été conçu dès le début pour imiter la démarche des humains dans la résolution des problèmes. À l'intérieur de la classe limitée d'énigmes qu'il pouvait traiter, l'ordre selon lequel il considérait les sous-buts et les actions possibles s'est montré comparable à celui des humains abordant les mêmes problèmes. C'est ainsi que GPS a certainement été le premier programme à intégrer l'approche de la « pensée humaine ». Le succès comme modèles cognitifs de GPS et des programmes qui ont suivi a conduit Newell et Simon (1976) à formuler la célèbre hypothèse du **système symbolique matériel** selon laquelle « un système symbolique matériel contient les moyens nécessaires et suffisants pour un comportement généralement intelligent ». Ils entendaient par là que tout système faisant preuve d'intelligence doit opérer en manipulant des structures de données composées de symboles. On verra que cette hypothèse a été contestée de plusieurs manières.

Chez IBM, Nathaniel Rochester et ses collègues ont produit plusieurs des premiers programmes d'IA. Herbert Gelernter (1959) a construit le *Geometry Theorem Prover*, capable de démontrer des théorèmes que de nombreux étudiants trouveraient difficiles. À partir de 1952, Arthur Samuel a écrit une série de programmes pour jouer aux dames, qui finirent par apprendre à jouer à un bon niveau d'amateur. Dans le même temps, il rendit caduque l'idée selon laquelle les ordinateurs ne pouvaient faire que ce qui leur était demandé : son

programme pouvait apprendre rapidement à jouer à un meilleur niveau que son créateur. Le programme fut présenté à la télévision en février 1956 et fit une forte impression. Comme Turing, Samuel avait du mal à obtenir du temps de calcul. Travaillant la nuit, il utilisait des machines restées à l'étage des tests dans l'usine d'IBM. Le chapitre 5 est consacré aux jeux tandis que le chapitre 21 explique les techniques d'apprentissage employées par Samuel.

John McCarthy déménagea de Dartmouth au MIT pour y effectuer trois contributions cruciales au cours d'une année historique : 1958. Dans le mémo n° 1 du laboratoire d'IA, McCarthy définit le langage **Lisp** qui allait devenir le langage de programmation dominant en IA pour les trente années à venir. Avec Lisp, McCarthy disposait de l'outil dont il avait besoin, mais il était confronté au problème de l'accès à des ressources informatiques rares et onéreuses. Pour y remédier, il inventa, avec d'autres collègues du MIT, le « temps partagé ». C'est aussi en 1958 que McCarthy a publié l'article intitulé « Programs with common sense », dans lequel il décrit *Advice Taker*, un programme hypothétique qui peut être considéré comme le premier système complet d'IA. Comme le *Logic Theorist* et le *Geometry Theorem Prover*, le programme de McCarthy était conçu afin d'utiliser des connaissances pour rechercher des solutions à des problèmes. Mais, à la différence des autres, il pouvait intégrer une connaissance générale du monde. Par exemple, son auteur montra comment, grâce à des axiomes simples, le programme pouvait générer un plan pour aller à l'aéroport. Le programme était également conçu pour pouvoir accepter de nouveaux axiomes dans le cours normal des opérations, ce qui lui donnait la possibilité d'acquérir des compétences dans de nouveaux domaines *sans reprogrammation préalable. Advice Taker* incorporait donc les principes de base de la représentation des connaissances et du raisonnement, à savoir qu'il est utile de posséder une représentation formelle et explicite du monde et de ses rouages, ainsi que de pouvoir manipuler cette représentation à l'aide de processus déductifs. Il est remarquable de constater à quel point l'essentiel de cet article de 1958 demeure encore actuel.

C'est aussi en 1958 que Marvin Minsky s'installa au MIT. Toutefois, sa collaboration avec McCarthy ne dura guère. McCarthy accordait beaucoup d'importance à la logique formelle dans les représentations et les raisonnements tandis que Minsky cherchait surtout à créer des programmes qui fonctionnent correctement et adoptait le cas échéant une approche antilogique. En 1963, McCarthy créa le laboratoire d'IA de Stanford. Son projet de recourir à la logique pour construire l'*Advice Taker* définitif profita de la découverte réalisée en 1965 par J. A. Robinson de la méthode de résolution (un algorithme de démonstration des théorèmes de la logique du premier ordre ; voir chapitre 9). Les travaux réalisés à Stanford insistaient sur l'emploi de méthodes générales de raisonnement logique. Parmi les applications de la logique prises en compte figuraient les systèmes de réponses à des questions et de planification de Cordell Green (1969b), ainsi que le projet de robotique Shakey du SRI (*Stanford Research Institute*). Ce projet présenté au chapitre 25 était le premier à intégrer complètement le raisonnement logique et l'activité physique.

Les étudiants supervisés par Minsky avaient choisi des problèmes limités, dont la solution exigeait apparemment le recours à l'intelligence. Ces domaines limités sont, depuis, connus sous le nom de **micromondes**. Le programme SAINT (1963), de James Slagle, pouvait résoudre des problèmes d'intégration en « forme close » typiques de ceux étudiés dans les classes préparatoires scientifiques. Le programme ANALOGY (1968), de Tom Evans, résolvait des problèmes d'analogies géométriques tels ceux des tests de QI. De son côté, le programme STUDENT, de Daniel Bobrow (1967) résolvait des problèmes d'algèbre élémentaire, comme :

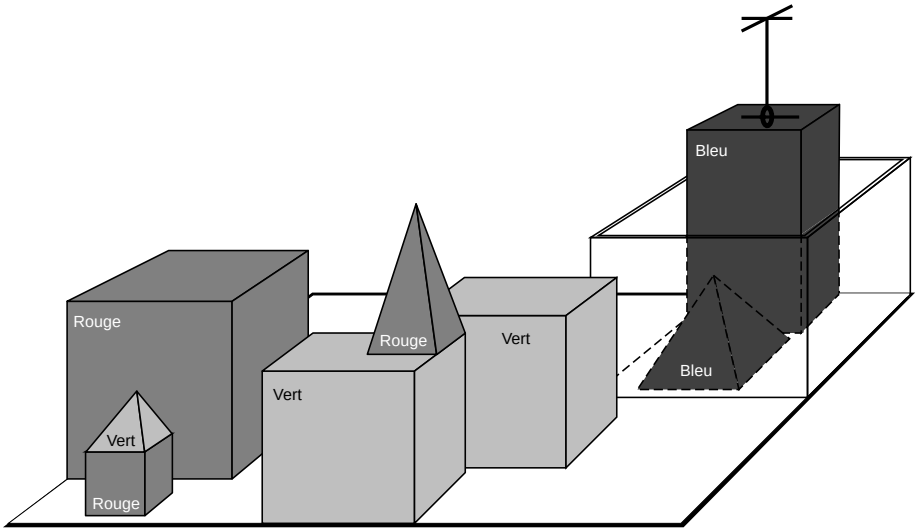

Figure 1.4 : Une scène du monde des blocs. SHRDLU (Winograd, 1972) vient de réaliser la commande « Trouve un bloc plus grand que celui que tu tiens et mets-le dans la boîte. »

> Si le nombre de clients obtenus par Tom est le carré de 20 % du nombre d'annonces qu'il a publiées et que le nombre de ces annonces est 45, combien Tom a-t-il gagné de clients ?

Le plus célèbre des micromondes fut le monde des blocs : un ensemble de blocs placés sur une table (ou plus souvent une simulation de table, voir figure 1.4). Dans ce monde, une tâche typique consiste à changer la disposition des blocs à l'aide d'un robot doté d'une pince qui peut saisir un bloc à la fois. Le monde des blocs a donné lieu au développement du projet de vision de David Huffman (1971), du travail sur la vision et la propagation des contraintes de David Waltz (1975), de la théorie de l'apprentissage de Patrick Winston (1970), du programme de compréhension du langage naturel de Terry Winograd (1972) et du planificateur de Scott Fahlman (1974).

Les premiers travaux de McCulloch et de Pitts sur les réseaux de neurones ont aussi commencé à porter leurs fruits. Le travail de Winograd et de Cowan (1963) a montré comment un grand nombre d'éléments pouvaient collectivement représenter un concept individuel, avec une augmentation correspondante de la robustesse et du parallélisme. Les méthodes d'apprentissage de Hebb ont été améliorées par Bernie Widrow (Widrow et Hoff, 1960 ; Widrow, 1962), qui a appelé ses réseaux **adalines**, et par Frank Rosenblatt avec ses **perceptrons** (1962). Le **théorème de convergence du perceptron**(Block *et al.*, 1962) énonce que l'algorithme d'apprentissage peut ajuster les intensités des liens d'un perceptron afin de les corréler à n'importe quelle donnée d'entrée dès lors qu'une telle corrélation existe. Ces sujets sont traités au chapitre 20.

1.3.4 L'épreuve de la réalité (1966 – 1973)

Dès le début, les chercheurs en IA n'ont pas craint de faire des prédictions quant à leurs succès futurs. On cite souvent le passage suivant, écrit en 1957 par Herbert Simon :

> Mon intention n'est pas de vous surprendre ou de vous choquer, mais la manière la plus simple de résumer les choses consiste à dire qu'il existe désormais des machines capables

> de penser, d'apprendre et de créer. En outre, leur capacité d'accomplir ces choses va rapidement s'accroître jusqu'à ce que, dans un futur proche, le champ des problèmes qu'elles pourront aborder soit coextensif à celui auquel s'applique l'esprit humain.

Une expression telle que « dans un futur proche » peut recevoir des interprétations très diverses, mais Simon faisait aussi des prédictions plus concrètes : dans les dix années à venir, un ordinateur serait champion d'échecs et un théorème mathématique important serait démontré par une machine. Il a fallu quarante ans au lieu de dix pour que ces prédictions se réalisent (ou presque). L'excès de confiance dont Simon faisait preuve était dû aux performances prometteuses des premiers systèmes d'IA sur des exemples simples. Mais, dans presque tous les cas, ces premiers systèmes ont lamentablement échoué lorsqu'ils ont été confrontés à des problèmes de plus grande envergure ou plus complexes.

La première difficulté venait de ce que les premiers programmes n'avaient aucune connaissance relative au sujet du problème : ils réussissaient en recourant uniquement à de simples manipulations syntaxiques. Un exemple typique est arrivé pendant les premiers efforts en traduction automatique : l'US National Research Council les avait généreusement financé afin d'accélérer la traduction de documents scientifiques russes à la veille du lancement du Spoutnik en 1957. On pensait au début que de simples transformations syntaxiques fondées sur les grammaires du russe et de l'anglais ainsi que le remplacement des mots à l'aide de dictionnaires électroniques suffiraient à restituer la signification exacte des phrases. Mais on sait maintenant qu'une connaissance de fond du sujet d'un texte est nécessaire à une traduction précise, pour permettre de résoudre les ambiguïtés éventuelles et de déterminer le sens des phrases à traduire : la fameuse retraduction de la phrase « l'esprit est fort mais la chair est faible » en « la vodka est bonne mais la viande est avariée » illustre les difficultés rencontrées. En 1966, le rapport d'un comité consultatif indiquait qu'il n'existait pas de machines capables de traduire des textes scientifiques généraux et qu'il était impossible d'en envisager une dans l'immédiat. Toutes les subventions fédérales pour les projets universitaires portant sur la traduction furent annulées. De nos jours, les traducteurs automatiques sont des outils imparfaits, même s'ils sont largement utilisés pour les documents techniques, commerciaux, administratifs et sur Internet.

Le deuxième type de problème était lié à l'impraticabilité des nombreux problèmes que l'IA essayait de résoudre. La plupart des programmes résolvaient des problèmes en essayant différentes combinaisons jusqu'à atteindre la solution. Cette stratégie convenait au début parce que les micromondes contenaient très peu d'objets, et donc très peu d'actions possibles et très peu de solutions. Avant le développement de la théorie de la complexité calculatoire, on s'accordait à penser qu'il suffirait de disposer de matériels plus rapides et d'une plus grande capacité de mémoire pour pouvoir aborder des problèmes de plus grande taille. L'optimisme suscité par le développement de la démonstration automatique de théorèmes par résolution a disparu sitôt que l'impossibilité de prouver des théorèmes mettant en jeu plus de quelques dizaines de faits est devenue patente. Les chercheurs venaient de comprendre que *le fait qu'un programme puisse trouver une solution en principe ne signifie pas qu'il contienne des mécanismes lui permettant de la trouver en pratique.*

L'illusion de la puissance de calcul illimitée n'était pas confinée aux seuls programmes résolvant des problèmes. Les premières expériences dans le domaine des **algorithmes génétiques** (on parlait à l'époque d'**évolution artificielle**) (Friedberg, 1958 ; Friedberg *et al.*, 1959) reposaient sur la croyance indéniablement correcte selon laquelle l'application d'une série appropriée de petites mutations à un programme écrit en code machine permettrait

de générer un programme caractérisé par de bonnes performances sur des tâches simples. L'idée de base consistait à essayer de manière aléatoire diverses mutations associées à un processus de sélection qui préserverait les variations les plus utiles en apparence, mais, malgré des milliers d'heures de temps machine, aucun progrès ou presque n'avait été enregistré. Les algorithmes génétiques modernes s'appuient sur de meilleures représentations et obtiennent davantage de succès.

Les rédacteurs du rapport Lighthill (Lighthill, 1973) reprochèrent surtout à l'IA de ne pas remédier à l'« explosion combinatoire » : c'est au vu de ce document que le gouvernement britannique décida de ne plus subventionner la recherche en IA que dans deux universités. (La tradition orale brosse un tableau différent et plus coloré : elle dépeint des ambitions politiques et des animosités personnelles qu'il ne saurait être question de décrire ici.)

Une troisième difficulté tenait aux restrictions fondamentales relatives aux structures de base susceptibles de générer un comportement intelligent. Le livre *Perceptrons* (1969) de Minsky et Papert démontra par exemple que si les perceptrons (une forme simplifiée des réseaux de neurones) pouvaient apprendre tout ce qu'ils parvenaient à représenter, leur capacité de représentation était limitée. C'est pourquoi un perceptron à deux entrées (simplifié par rapport à la forme que Rosenblatt avait étudié à l'origine) ne pouvait pas apprendre à reconnaître les cas où ses deux entrées étaient différentes. Même si les conclusions de ces auteurs ne s'appliquaient pas aux réseaux plus complexes ou multicouches, le financement des recherches sur les réseaux de neurones s'amenuisa au point de se réduire à presque rien. L'ironie du sort a voulu que les nouveaux algorithmes d'apprentissage par rétropropagation pour les réseaux multicouches, qui ont relancé les recherches sur les réseaux de neurones à la fin des années 1980, aient été découverts en 1969 (Bryson et Ho, 1969).

1.3.5 Systèmes fondés sur les connaissances (1969 – 1979) : la clé de la puissance ?

Le paradigme de résolution de problèmes élaboré au cours de la première décennie de recherche en IA consistait en un mécanisme de recherche d'ordre général qui essayait d'enchaîner des étapes de raisonnement élémentaires pour trouver des solutions complètes. De telles approches ont été qualifiées de **méthodes faibles** car, quoique générales, elles ne supportent pas le changement d'échelle pour résoudre des problèmes plus grands ou plus difficiles. L'alternative aux méthodes faibles est de recourir à des connaissances plus puissantes et spécifiques au domaine concerné, qui permettent des étapes de raisonnement plus importantes et gèrent plus facilement les cas typiques rencontrés dans des domaines d'expertise limités. On pourrait dire que, pour résoudre un problème difficile, il est presque obligatoire d'en connaître la solution à l'avance.

Le programme DENDRAL (Buchanan *et al.*, 1969) constitua l'un des premiers exemples de cette approche. Il a été développé à Stanford, où Ed Feigenbaum (ancien étudiant de Herbert Simon), Bruce Buchanan (philosophe devenu chercheur en informatique) et Joshua Lederberg (généticien lauréat du prix Nobel) ont conjointement résolu le problème de l'inférence d'une structure moléculaire à partir des informations fournies par un spectromètre de masse. L'entrée du programme consiste en une formule élémentaire de la molécule (par exemple, $C_6H_{13}NO_2$ et du spectre de masse qui donne les masses des différents fragments de la molécule générée lorsqu'elle est bombardée par un faisceau d'électrons. Par exemple, le spectre de masse est susceptible de contenir une pointe à $m = 15$, soit la masse d'un fragment de méthyle (CH_3).

La version naïve du programme générait toutes les structures composables à partir de la formule de départ et prédisait ensuite le spectre de masse observable pour chacune, après quoi il comparait cette donnée au spectre effectif. Comme on pouvait s'y attendre, il apparut que cette procédure était impraticable même pour des molécules d'une taille modérée : des experts en chimie analytique consultés par les chercheurs de DENDRAL expliquèrent qu'ils avaient pour habitude de rechercher dans les spectres des profils en pic bien connus qui suggéraient l'existence de sous-structures communes dans la molécule. Par exemple, la règle suivante sert à reconnaître un groupement carbonyle (C=O, dont le poids est 28).

> **si** il y a deux pics à x_1 et x_2 tels que :
> (a) $x_1 + x_2 = M + 28$ (M correspondant à la masse totale de la molécule) ;
> (b) $x_1 - 28$ est un pic élevé ;
> (c) $x_2 - 28$ est un pic élevé ;
> (d) au moins une des deux valeurs x_1 et x_2 est élevée ;
> **alors** il y a un groupement carbonyle.

Le fait de détecter la présence d'une sous-structure particulière dans une molécule réduit considérablement le nombre de candidats possibles. La puissance de DENDRAL provenait de ce que :

> Toute la connaissance théorique pertinente pour la résolution de ces problèmes avait été formalisée dans une table établissant une correspondance entre la forme générale du [spectre attendu du composant] (« premiers principes ») et les formes spéciales pertinentes (« les recettes de cuisine ») (Feigenbaum *et al.*, 1971).

DENDRAL était important parce que c'était le premier système qui réussissait à faire un usage *intensif* des *connaissances* : son expertise dérivait d'un grand nombre de règles spécialisées. Les systèmes ultérieurs ont intégré le thème principal de l'approche adoptée par McCarthy pour *Advice Taker* – une séparation nette entre les connaissances (sous la forme de règles) et le composant de raisonnement.

C'est en gardant cet enseignement à l'esprit que Feigenbaum et ses collègues de Stanford lancèrent le projet HPP (*Heuristic Programming Project*) afin de déterminer les possibilités d'application des **systèmes experts** à d'autres domaines d'expertise. L'essentiel de l'effort suivant fut orienté vers le domaine du diagnostic médical. Feigenbaum, Buchanan et le Dr Edward Shortliffe développèrent MYCIN pour diagnostiquer des affections sanguines. Avec 450 règles environ, MYCIN présentait les mêmes performances que certains experts et obtenait des résultats bien meilleurs que des médecins fraîchement diplômés. Ce système différait de DENDRAL sur deux points importants. En premier lieu, aucun modèle théorique existant n'aurait pu permettre à MYCIN de déduire des règles : il fallait les acquérir en s'entretenant longuement avec des experts, qui les avaient eux-mêmes acquises dans des manuels, auprès d'autres experts et *via* leur expérience directe. En second lieu, les règles devaient refléter l'incertitude inhérente au savoir médical : MYCIN intégrait un calcul de l'incertitude fondé sur des **facteurs de certitude** (voir chapitre 14), qui paraissait (à l'époque) bien correspondre à la manière dont les médecins évaluent l'impact de l'observation sur le diagnostic.

L'importance de la prise en compte des connaissances était également manifeste dans le domaine de la compréhension du langage naturel. Bien que le système SHRDLU de Winograd ait suscité beaucoup d'intérêt pour ce domaine, sa dépendance à l'égard de l'analyse syntaxique posait les mêmes problèmes que ceux rencontrés par les premières machines traductrices. Il pouvait remédier aux ambiguïtés et comprendre les références pronominales, mais ces caractéristiques étaient surtout liées au fait qu'il avait été conçu pour un domaine

spécifique : le monde des blocs. Selon plusieurs chercheurs – tel Eugene Charniak, collègue de Winograd au MIT – toute compréhension solide du langage suppose une connaissance générale du monde et une méthode générique d'utilisation de cette connaissance.

À Yale, Roger Schank, linguiste devenu chercheur en IA, insista sur ce point en déclarant que « la syntaxe n'existe pas », ce qui provoqua un tollé chez de nombreux experts en linguistique mais permit de lancer une discussion fructueuse. Schank et ses étudiants construisirent une série de programmes (Schank et Abelson, 1977 ; Wilensky, 1978 ; Schank et Riesbeck, 1981 ; Dyer, 1983) dont la tâche était de comprendre le langage naturel. Cependant, l'accent portait moins sur le langage *en soi* que sur les problèmes de représentation des connaissances nécessaires à la compréhension du langage et les problèmes de raisonnement associés, notamment la représentation de situations stéréotypées (Cullingford, 1981), la description de l'organisation de la mémoire chez les humains(Rieger, 1976 ; Kolodner, 1983) et la compréhension de plans et de buts (Wilensky, 1983).

L'apparition généralisée d'applications aux problèmes du monde réel provoqua une augmentation concomitante de la demande de schémas de représentation des connaissances opérationnels. De nombreux langages de représentation et de raisonnement furent développés. Certains étaient fondés sur la logique : c'est le cas de Prolog – langage très apprécié en Europe – et de la famille de langages PLANNER aux États-Unis. D'autres, fondés sur le concept des **schémas** (*frames*) développé par Minsky (1975), ont adopté une approche plus structurée : ils assemblent des observations relatives à des objets particuliers et à des types d'événements, puis ordonnent les types dans une vaste hiérarchie taxonomique analogue à celle de la biologie.

1.3.6 L'IA devient une industrie (de 1980 à nos jours)

Le premier système expert commercial réussi, R1, est né chez Digital Equipment Corporation (McDermott, 1982). Ce programme configurait des ordinateurs en fonction des commandes des clients ; en 1986, on estimait qu'il avait permis à l'entreprise d'économiser près de 40 millions de dollars par an. En 1988, le département IA de DEC avait déjà déployé 40 systèmes experts et d'autres étaient prévus. DuPont en utilisait 100 et en développait 500, ce qui lui donnait la possibilité d'économiser environ 10 millions de dollars par an. Presque toutes les grandes entreprises des États-Unis possédaient un département IA et utilisaient des systèmes experts ou envisageaient d'en utiliser.

En 1981, les Japonais annoncèrent le lancement du projet Cinquième génération, plan décennal qui prévoyait de construire des ordinateurs intelligents programmés en Prolog. En réponse à cette annonce, les États-Unis créèrent un consortium de recherche nommé MCC (*Microelectronics and Computer Technology Corporation*) qui visait à préserver la compétitivité américaine en ce domaine. Dans les deux cas, l'IA faisait partie d'un effort plus vaste, comprenant des recherches sur les architectures de puces et sur les interfaces homme-machine. En Grande-Bretagne, le rapport Alvey rétablit les financements supprimés à la suite du rapport Lighthill [14]. Les projets n'ont jamais atteint leurs ambitieux objectifs dans ces trois pays.

Dans l'ensemble, l'industrie de l'IA s'est vigoureusement développée, son chiffre d'affaires passant de quelques millions de dollars en 1980 à plusieurs milliards de dollars en 1988, avec des centaines d'entreprises qui vendaient des systèmes experts, des systèmes de vision

14. Pour sauver les apparences, on inventa un nouveau domaine appelé IKBS (*Intelligent Knowledge-Based Systems*), car il ne pouvait être officiellement question d'intelligence artificielle.

artificielle, des robots, ainsi que du matériel et du logiciel spécialisé pour ces applications. Puis vint peu après une période de stagnation, dite « hiver de l'IA », au cours de laquelle de nombreuses entreprises ont abandonné à cause de leur incapacité à tenir des promesses extravagantes.

1.3.7 Retour des réseaux de neurones (de 1986 à nos jours)

Au milieu des années 1980, quatre groupes de chercheurs au moins réinventèrent l'algorithme d'apprentissage par **rétropropagation**, initialement mis au point par Bryson et Ho en 1969 (voir chapitre 20). Cet algorithme fut appliqué à de nombreux problèmes d'apprentissage en informatique et en psychologie, et la publication des résultats dans la collection « Parallel Distributed Processing » (Rumelhart et McClelland, 1986) suscita énormément d'enthousiasme.

Certains estimaient que les modèles de systèmes intelligents dits **connexionnistes** concurrenceraient directement tant les modèles symboliques prônés par Newell et Simon que l'approche logiciste de McCarthy et d'autres chercheurs (Smolensky, 1988). Il peut sembler évident que les humains manipulent des symboles à un certain niveau – l'ouvrage de Terrence Deacon, *The Symbolic Species* (1997), va même jusqu'à préciser qu'il s'agirait là d'un *trait distinctif* du genre humain, mais les connexionnistes les plus ardents remettaient en cause le rôle explicatif de la manipulation des symboles dans des modèles détaillés de la cognition. Si cette question demeure non résolue, l'opinion la plus largement admise est que les approches connexionniste et symbolique sont complémentaires et non pas concurrentes. De la même façon que l'IA et les sciences cognitives se sont séparées, la recherche moderne sur les réseaux de neurones s'est scindée en deux domaines, l'un concernant la construction d'architectures et d'algorithmes neuronaux efficaces, ainsi que la compréhension de leurs propriétés mathématiques, l'autre concernant la modélisation précise des propriétés empiriques des vrais neurones ou ensembles de neurones.

1.3.8 L'IA devient une science (de 1987 à nos jours)

Ces dernières années ont été le théâtre d'une révolution, tant au niveau des thèmes de recherche que des méthodologies adoptées en intelligence artificielle [15]. Il est désormais plus fréquent de reprendre des théories existantes que d'en proposer de nouvelles, de fonder des propositions sur des théorèmes rigoureux ou des faits expérimentaux plutôt que sur l'intuition et de préférer les applications du monde réel aux exemples de laboratoire. Les origines de l'IA tiennent en partie à une rébellion contre les limites des domaines tels que la théorie du contrôle et les statistiques, domaines qui furent ensuite absorbés. Comme David McAllester (1998) l'a écrit :

> Au cours de la première période de l'IA, tout portait à croire que de nouvelles formes de calcul symbolique, par exemple les schémas et les réseaux sémantiques, allaient largement contribuer à l'obsolescence des théories classiques. Cela déboucha sur une sorte d'isolationnisme qui sépara clairement l'IA du reste de l'informatique : c'est cet isolationnisme qui est actuellement abandonné. On reconnaît que l'apprentissage artificiel ne doit pas

15. Certains y ont vu une victoire des ***neats*** (ceux qui pensent que les théories de l'IA doivent être mathématiquement rigoureuses) sur les ***scruffies*** (ceux qui, tenant à essayer toutes sortes d'idées, préfèrent écrire des programmes puis voir ce qui semble marcher), mais ces deux approches sont aussi importantes l'une que l'autre. Un virage en direction de la « propreté » indique que le domaine est devenu stable et mature – que cette stabilité soit un jour remise en cause par une nouvelle idée « débraillée » est une autre question.

> être isolé de la théorie de l'information, que le raisonnement en environnement incertain ne doit pas être dissocié des modélisations stochastiques, que la recherche ne doit pas être séparée de l'optimisation et du contrôle classiques, et que le raisonnement automatique ne doit pas être disjoint des méthodes formelles et de l'analyse statique.

Du point de vue de la méthodologie, l'IA s'est de plus en plus conformée à l'approche scientifique : on a fini par comprendre que des hypothèses ne peuvent être acceptées qu'après des expérimentations rigoureuses et que l'importance des résultats obtenus doit être évaluée par analyse statistique (Cohen, 1995). Il est désormais possible de reproduire des expérimentations grâce aux dépôts de code et de données de test.

Le domaine de la reconnaissance de la parole illustre ce schéma. Dans les années 1970, une grande diversité d'architectures et d'approches a été tentée. Nombre d'entre elles n'étaient que des constructions *ad hoc* et fragiles qui ne reposaient que sur quelques exemples sélectionnés arbitrairement. Depuis peu, des approches fondées sur les **modèles de Markov cachés** (MMC, ou HMM, *Hidden Markov Models*) dominent cette discipline, et deux caractéristiques des MMC l'expliquent. Premièrement, ils s'étayent sur une théorie mathématique rigoureuse, ce qui a permis aux chercheurs en reconnaissance de la parole de s'appuyer sur plusieurs décennies de résultats mathématiques acquis dans d'autres domaines. Deuxièmement, ils sont générés par un processus d'apprentissage mené sur un corpus important de données réelles, ce qui assure la robustesse des performances : des tests rigoureux effectués en aveugle sur des modèles MMC ont montré que ceux-ci améliorent constamment leurs scores. Cette technologie, comme le domaine voisin de la reconnaissance des caractères, a déjà effectué la transition vers des applications grand public et industrielles. Attention : personne ne peut scientifiquement avancer que les humains utilisent des MMC pour la reconnaissance de la parole. En revanche, les MMC fournissent un cadre mathématique pour comprendre le problème et affirmer, du point de vue de l'ingénierie, qu'ils fonctionnent bien en pratique.

La traduction automatique suit le même parcours que la reconnaissance de la parole. Dans les années 1950, il y a eu une grande fébrilité autour d'une approche fondée sur les séquences de mots, avec des modèles appris selon les lois de la théorie de l'information. Cette approche, tombée dans l'oubli dans les années 1960, mais est revenue sur le devant de la scène à la fin des années 1990 et domine maintenant le domaine.

Les réseaux de neurones se conforment aussi à cette tendance. Les travaux conduits en ce domaine au cours des années 1980 aspiraient surtout à déterminer ce qui pouvait être concrètement réalisé et à comprendre en quoi les réseaux de neurones diffèrent des techniques « traditionnelles ». Grâce à une méthodologie et à un cadre théorique améliorés, les réseaux de neurones ont atteint un stade qui autorise désormais à les comparer aux techniques correspondantes de la statistique, de la reconnaissance des formes et de l'apprentissage artificiel, ce qui permet d'employer pour chaque application la technique la plus prometteuse. Le **data mining** (ou **fouille de données**) est une technologie résultant de ces développements qui a donné lieu à l'émergence d'une nouvelle industrie des plus dynamiques.

Après l'article de Peter Cheeseman « In defense of probability » (1985), l'ouvrage de Judea Pearl intitulé *Probabilistic Reasoning in Intelligent Systems* (1988) a conduit à une nouvelle acceptation de la théorie des probabilités et de la décision en IA. Le formalisme des **réseaux bayésiens** a été inventé pour permettre de représenter efficacement des connaissances incertaines et de raisonner rigoureusement dans ce contexte. Cette approche qui remédie à de nombreux problèmes sur lesquels butaient les systèmes de raisonnement probabiliste des années 1960 et 1970 domine à présent la recherche sur le raisonnement en environnement

incertain et les systèmes experts : autorisant l'apprentissage à partir de l'expérience, elle combine le meilleur de l'IA classique et des réseaux de neurones. Les travaux de Judea Pearl (1982a), d'Eric Horvitz et de David Heckerman (Horvitz et Heckerman, 1986 ; Horvitz *et al.*, 1986) ont développé la notion de systèmes experts *normatifs :* des systèmes qui agissent rationnellement conformément aux lois de la théorie de la décision et n'essaient pas d'imiter les étapes de raisonnement des experts humains. Le système d'exploitation Windows^{TM} contient plusieurs systèmes experts de diagnostic de ce type (voir chapitres 13 à 16).

Des révolutions du même ordre se sont produites en robotique, en vision informatique et en représentation des connaissances. Une meilleure compréhension des problèmes et de leurs propriétés de complexité, conjuguée à une sophistication accrue des outils mathématiques, a permis d'élaborer des programmes de recherches réalisables et des méthodes fiables. Bien qu'un acroissement de la formalisation et de la spécialisation a conduit des domaines comme la vision artificielle et la robotique à s'isoler du « courant principal » de l'IA dans les années 1990, cette tendance s'est renversée ces dernières années, grâce aux outils de l'apprentissage artificiel qui ont prouvé leur efficacité sur de nombreux problèmes. Le processus de réintégration est déjà très profitable.

1.3.9 Émergence des agents intelligents (de 1995 à nos jours)

Peut-être encouragés par les progrès accomplis dans la résolution des sous-problèmes de l'IA, des chercheurs ont également reconsidéré le problème de l'« agent total ». Les travaux d'Allen Newell, de John Laird et de Paul Rosenbloom sur SOAR (Newell, 1990 ; Laird *et al.*, 1987) constituent l'exemple le plus connu d'architecture d'agent complet. Le mouvement dit « situé » a pour but de comprendre le fonctionnement d'agents immergés dans des environnements réels et exposés à des entrées sensorielles continues.

Son objectif est la compréhension des mécanismes des agents immergés dans des environnements réels et soumis à des perceptions continues. Internet fait partie des environnements les plus importants pour les agents intelligents. Les systèmes d'IA sont devenus si courants dans les applications web que le suffixe « bot » a fini par entrer dans le langage courant. De plus, de nombreux outils Internet tels que les moteurs de recherche, les systèmes de recommandation et les agrégateurs de sites web, reposent sur des techniques d'IA.

Ces tentatives de construire des agents complets ont permis de constater que les sous-domaines précédemment isolés de l'IA ont d'autant plus besoin d'être réorganisés que leurs résultats sont destinés à se compléter. En particulier, il est à présent largement admis que les systèmes sensoriels (vision, sonar, reconnaissance vocale, etc.) ne peuvent pas fournir d'informations environnementales fiables : il faut donc que les systèmes de raisonnement et de planification puissent gérer l'incertitude. Autre conséquence importante de la perspective des agents : l'IA doit entretenir des relations plus étroites avec les autres domaines – notamment l'économie et la théorie du contrôle – dans lesquels les agents jouent également un rôle. Les avancées récentes dans le pilotage de véhicules robotisés ont été possibles grâce à la fusion de plusieurs approches, parmi lesquelles le développement de meilleurs capteurs, d'une intégration au niveau automatique de la perception, de la cartographie et la localisation et enfin d'une dose de planification à haut niveau.

En dépit de ces succès, plusieurs pères fondateurs de l'IA, parmi lesquels John McCarthy (2007), Marvin Minsky (2007), Nils Nilsson (1995, 2005) et Patrick Winston (Beal et Winston, 2009) ont exprimé leur mécontentement par rapport aux développements récents de l'IA. Ils pensent que l'IA devrait moins se focaliser sur la création d'applications toujours plus

performantes sur des tâches très spécifiques, telles que conduire une voiture, jouer aux échecs ou reconnaître de la parole. Au lieu de ça, ils croient que l'IA devrait revenir à ses racines et se battre pour, selon les propres mots de Simon, « des machines qui pensent, qui apprennent et qui créent ». Ils appellent cette tendance **IA de niveau humain** ou IANH (HLAI, « Human-Level Artificial Intelligence ») ; leur premier symposium a eu lieu en 2004 (Minsky *et al.*, 2004). Cet effort demandera de très grandes bases de données ; on peut trouver des idées sur des sources possibles pour ces bases de données dans Hendler *et al.* (1995).

Le sous-domaine de l'**intelligence artificielle générale** ou IAG (Goertzel et Pennachin, 2007) est une idée connexe ; la première conférence a eu lieu en 2008 et a été organisée par le *Journal of Artificial General Intelligence*. L'IAG cherche un algorithme universel pour apprendre et agir dans n'importe quel environnement, et trouve ses racines dans les travaux de Ray Solomonoff (1964), l'un des participants à la conférence originelle de Dartmouth en 1956. La création de quelque chose qui soit vraiment une **IA amicale** est également une préoccupation (Yudkowsky, 2008 ; Omohundro, 2008) sur laquelle nous reviendrons au chapitre 26.

1.3.10 La disponibilité de vastes ensembles de données (de 2001 à nos jours)

Pendant les soixante ans de l'histoire de l'informatique, l'accent a été mis sur *l'algorithme* comme objet central d'étude. Des travaux récents en IA suggèrent que pour de nombreux problèmes, il serait plus sensé de s'intéresser aux *données* et d'être moins exigeant sur le choix de l'algorithme à leur appliquer. C'est un point qui se défend parce qu'on dispose de sources de données de plus en plus grandes : par exemple, des milliers de milliards d'occurrences de mots en anglais et des milliards d'images sur le Web (Kilgarriff et Grefenstette, 2006) ; ou encore des milliards de paires de bases de séquences de gènes (Collins *et al.*, 2003).

Dans cette direction, le travail de Yarowsky (1995) sur la désambiguïsation des mots a été séminal : étant donné l'utilisation du mot « plant » (NDT : en anglais) dans une phrase, s'agit-il d'une plante ou d'une usine ? Les approches précédentes reposaient sur des exemples étiquetés à la main combinés avec des algorithmes d'apprentissage artificiel. Yarowsky a montré que cette tâche pouvait être réalisée avec une précision supérieure à 96 % sans exemples étiquetés. À la place, étant donné un très grand corpus et les seules définitions des deux sens du mot – usine et végétal – on peut étiqueter les exemples du corpus, et partant de là, **amorcer** l'apprentissage de motifs qui aideront à étiqueter de nouveaux exemples. Banko et Brill (2001) montrent que ce genre de technique donne encore de meilleurs résultats quand la quantité de texte disponible passe d'un million de mots à un milliard, et que la qualité qu'on retire d'un plus gros corpus dépasse de loin celle qu'on retire à choisir un meilleur algorithme ; un algorithme médiocre entraîné avec 100 millions de mots non étiquetés enfonce le meilleur algorithme connu entraîné avec un million de mots.

Comme autre exemple, Hays et Efros (2007) abordent le problème de combler des trous dans des photographies. Supposons que vous utilisiez Photoshop pour gommer un ex-ami d'une photo de groupe. Ensuite, vous avez besoin de remplir la zone gommée avec quelque chose qui corresponde à l'arrière-plan. Hays et Efros ont conçu un algorithme qui recherche le motif dont vous avez besoin dans une collection de photos. Ils ont montré que la performance de leur algorithme était mauvaise quand ils utilisaient une collection de seulement dix mille photos mais devenait excellente quand ils passaient à deux millions de photos.

Des travaux comme ceux-ci laissent à penser que la « pierre d'achoppement » de la connaissance en IA – comment exprimer toute la connaissance dont un système a besoin – pourrait

être résolue dans beaucoup de cas par apprentissage plutôt que par codage manuel de la connaissance en question, à condition que les algorithmes d'apprentissage disposent de suffisamment de données d'entraînement (Halevy *et al.*, 2009). Des observateurs du domaine ont noté les nouvelles applications qui ont surgi et ils ont rapporté que « l'hiver de l'IA » pourrait donner lieu à un nouveau printemps(Havenstein, 2005). Comme l'écrit Kurzweil (2005), « aujourd'hui, des milliers d'applications IA sont profondément enfouies dans l'infrastructure de chaque industrie ».

1.4 État de l'art

Quelles sont les possibilités actuelles de l'IA ? Il est difficile d'apporter une réponse concise à cette question, car elle renvoie à des activités multiples appartenant à de nombreux sous-domaines. Les paragraphes suivants présentent quelques exemples d'applications : d'autres seront évoquées au fil de ce livre.

Véhicules autonomes. Une voiture robotisée sans pilote baptisée STANLEY gagna la course de 220 km, ainsi que l'édition 2005 du DARPA Grand Challenge en roulant sur le terrain accidenté du désert du Mojave à une vitesse moyenne de 35 km/h,. STANLEY est un Volkswagen Touareg bourré de caméras de radars, de télémètres laser pour percevoir son environnement et de logiciels qui peuvent prendre le contrôle du volant, des freins et de l'accélérateur (Thrun, 2006). L'année suivante, BOSS, développé à CMU, a gagné le « Urban Challenge » en conduisant en totale sécurité dans le flot du trafic des rues d'une base de l'armée de l'air, en obéissant aux règles de la circulation et en évitant les piétons et les autres véhicules.

Reconnaissance de la parole. Un système automatique de reconnaissance de la parole et de gestion du dialogue peut tenir toute la conversation avec un voyageur qui appelle United Airlines pour réserver un vol.

Planification et programmation autonome. À une centaine de millions de kilomètres de la Terre, REMOTE AGENT, de la NASA, est devenu le premier programme de planification autonome embarqué. Il a servi à contrôler la programmation des opérations à bord d'un vaisseau spatial (Jonsson *et al.*, 2000). REMOTE AGENT générait des plans à partir des objectifs généraux indiqués par le sol et en surveillait l'exécution ; il détectait, diagnostiquait et résolvait les problèmes qui survenaient. MAPGEN (Al-Chang *et al.*, 2004) est un programme ultérieur qui a planifié les opérations quotidiennes des robots d'exploration martiens Rovers (astromobiles) de la NASA. MEXAR2 (Cesta *et al.*, 2007) a organisé la mission Mars Express, qu'il s'agisse de la logistique ou de l'aspect scientifique, pour le compte de l'Agence spatiale européenne en 2008.

Jeux. DEEP BLUE d'IBM est le premier ordinateur qui soit parvenu à vaincre le champion du monde d'échecs Garry Kasparov sur un score de 3,5 contre 2,5 (Goodman et Keene, 1997). Kasparov a déclaré qu'il avait eu l'impression qu'« un nouveau type d'intelligence » était à l'œuvre sur l'échiquier. Le magazine *Newsweek* écrivit à propos de ce match qu'il s'agissait de « la dernière résistance du cerveau », et la valeur boursière d'IBM s'accrut de 18 milliards de dollars. Des champions humains ont étudié la défaite de Kasparov et ont ainsi pu arracher quelques victoires dans les années suivantes, mais les matches les plus récents entre l'homme et l'ordinateur ont été gagnés par l'ordinateur haut la main.

Anti-spam. Tous les jours, des algorithmes classent des milliards de messages électroniques comme spams, épargnant à leur destinataire de devoir détruire ce qui correspondrait à 80 % à 90 % de tous leurs messages s'ils n'étaient pas marqués comme indésirables auparavant. Étant donné que les spammeurs font continuellement évoluer leurs stratégies, un programme

statique ne peut pas lutter, mais les algorithmes par apprentissage marchent bien (Sahami *et al.*, 1998 ; Goodman et Heckerman, 2004).

Planification logistique. Au cours de la crise du golfe Persique survenue en 1991, les forces armées des États-Unis ont déployé DART (*Dynamic Analysis and Replanning Tool*), outil d'analyse et de replanification dynamique (Cross et Walker, 1994) qui automatise la gestion de la logistique et de la programmation des transports. Ceux-ci mettaient en œuvre jusqu'à 50 000 véhicules, bateaux et soldats à la fois, et il fallait tenir compte également des points de départ, des destinations, des routes et de la résolution des conflits potentiels, entre autres paramètres. On a généré un plan en quelques heures grâce aux techniques de l'IA, alors que les anciennes méthodes auraient demandé plusieurs semaines. Selon la DARPA (*Defense Advanced Research Project Agency*), cette application à elle seule faisait plus que rembourser les sommes investies depuis trente ans dans l'IA.

Robotique. La société iRobot Corporation a vendu plus de deux millions d'exemplaires de Roomba, aspirateurs robots à usage domestique. La même société exporte également en Irak et en Afghanistan, le plus redoutable PackBot qui est utilisé pour identifier des substances dangereuses, déminer, ou localiser des snipers.

Traduction automatique. Un programme informatique traduit automatiquement de l'arabe à l'anglais, permettant à un locuteur anglais de lire le titre « Ardogan confirme que la Turquie n'accepterait aucune pression, en leur demandant instamment qu'ils reconnaissent Chypre. » Le programme utilise un modèle statistique construit à partir d'exemple de traductions arabe-anglais et d'exemples de textes anglais totalisant deux mille milliards d'occurrences de mots (Brants *et al.*, 2007). Aucun des informaticiens de l'équipe ne parlait arabe, mais ils parlaient le langage des statistiques et des algorithmes d'apprentissage.

Les exemples précédents ont permis de présenter quelques-uns des systèmes d'intelligence artificielle existant de nos jours. Il y est question non pas de magie ou de science-fiction mais plutôt de sciences, d'ingénierie et de mathématiques, domaines auxquels le présent ouvrage introduit.

Résumé

Ce chapitre a défini l'IA et tracé l'arrière-plan historique et culturel de son développement :

- L'IA peut être envisagée avec différents objectifs en tête. Les deux questions essentielles qu'il convient de se poser sont celles-ci : vous intéressez-vous plutôt à la pensée ou au comportement ? Voulez-vous prendre modèle sur les humains ou travailler à partir d'une norme idéale ?
- Dans ce livre, nous adoptons le point de vue selon lequel l'intelligence a principalement trait à l'**action rationnelle**. Dans l'idéal, un **agent intelligent** exécute la meilleure action possible compte tenu de la situation. C'est dans cette perspective que nous étudions le problème de la construction d'agents intelligents.
- Les philosophes (dès l'an 400 av. J.-C.) ont rendu l'IA concevable en supposant que l'esprit peut être considéré à certains égards comme une machine, qu'il opère sur des connaissances encodées dans un langage interne et que la pensée peut permettre de choisir les actions à entreprendre.
- Les mathématiciens ont fourni les outils nécessaires à la manipulation d'énoncés logiques ou probabilistes (selon leur degré de certitude). Ils ont également défini les bases du calcul et du raisonnement algorithmique.

– Les économistes ont formalisé le problème de la prise de décisions qui maximisent les gains prévisibles pour le décideur.
– Les neurobiologistes ont fait certaines découvertes sur le fonctionnement du cerveau, et ses similitudes et ses différences avec un ordinateur.
– Les psychologues ont adopté l'idée selon laquelle les humains et les animaux peuvent être vus comme des machines de traitement de l'information. Les linguistes ont montré que l'usage du langage s'insère dans ce modèle.
– Les informaticiens ont fourni les machines de plus en plus puissantes qui rendent possibles les applications de l'IA.
– La théorie du contrôle traite de la conception de dispositifs opérant de manière optimale à partir du feed-back fourni par l'environnement. À l'origine, les outils mathématiques utilisés par cette discipline étaient différents de ceux de l'IA, mais ces deux domaines sont en train de se rapprocher l'un de l'autre.
– L'histoire de l'IA se caractérise par des phases de succès et d'optimisme démesuré, d'une part, et des périodes de pessimisme et de restrictions budgétaires, d'autre part. On remarque aussi des cycles d'introduction de nouvelles approches et de redéfinition systématique des meilleures d'entre elles.
– Les avancées de l'IA se sont accélérées au cours de la dernière décennie du fait d'un plus grand usage de la méthode scientifique dans l'expérimentation et de la comparaison des approches.
– Les progrès récemment accomplis dans la compréhension des bases théoriques de l'intelligence sont allés de pair avec des améliorations des capacités des systèmes réels. Les sous-domaines de l'IA ont été mieux intégrés et l'IA a trouvé un terrain commun avec d'autres disciplines.

Notes bibliographiques et historiques

Herb Simon explore le statut méthodologique de l'IA dans *The Sciences of the Artificial* (1981). Cet ouvrage examine les domaines de recherche relatifs aux artefacts complexes et explique dans quelle mesure l'IA relève à la fois de la science et des mathématiques. Cohen (1995) donne un aperçu de la méthode expérimentale en IA.

Le test de Turing est traité par Shieber (1994), qui est très sévère à l'égard de son instanciation dans la compétition du prix Loebner (*Loebner Prize*) et par Ford et Hayes (1995), qui soutiennent que le test en lui-même n'est pas utile à l'IA. Bringsjord (2008) explore la notion de juge pour le test de Turing. Shieber (2004) et Epstein *et al.* (2008) recensent des essais sur le test de Turing. *Artificial Intelligence: The Very Idea*, par John Haugeland (1985), offre un compte-rendu accessible des problèmes pratiques et philosophiques de l'IA. Les anthologies par Webber et Nilsson (1981) et par Luger (1995) regroupent les publications les plus importantes des débuts de l'IA. L'*Encyclopedia of AI* (Shapiro, 1992) contient des articles sur presque tous les sujets de l'IA, tout comme Wikipédia. Ces articles constituent souvent de bons points de départ pour explorer les articles de recherche sur un sujet donné. Nils Nilsson (2009), l'un des pionniers du domaine, retrace une histoire exhaustive et avisée de l'IA.

Les travaux les plus récents sont publiés dans les comptes-rendus des principales conférences consacrées à l'IA : l'*International Joint Conference on AI* (IJCAI), la *European Conference on AI* (ECAI, annuelle) et la *National Conference on AI*, plus souvent appelée AAAI, du nom de l'organisme qui la sponsorise. Les principales publications généralistes en IA sont *Artificial Intelligence*, *Computational Intelligence*, *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, *IEEE Intelligent Systems* et le *Journal of Artificial Intelligence Research* (une publication électronique). On trouve aussi de nombreuses conférences et publications consacrées à des domaines spécifiques (voir les chapitres appropriés). Les principales associations professionnelles sont l'*American Association for Artificial Intelligence* (AAAI), l'*ACM Special Interest Group on Artificial Intelligence* (SIGART) et la *Society for Artificial Intelligence and Simulation of Behaviour* (AISB). L'*AI Magazine* (de l'IAAA) contient de nombreux articles et didacticiels ; vous trouverez sur son site web, `aaai.org`, des nouvelles et des informations d'ordre général.

Exercices

Les exercices qui suivent ont été conçus de manière à susciter des débats et certains d'entre eux peuvent être utilisés dans le cadre d'un projet plus long. On peut également les essayer dès à présent et revenir sur ces premiers essais après avoir achevé la lecture de ce livre.

1.1 Définissez dans vos propres termes : (a) intelligence, (b) intelligence artificielle, (c) agent, (d) rationalité, (e) raisonnement logique.

1.2 Lisez l'article de Turing sur l'IA (Turing, 1950). Cet article passe en revue plusieurs objections opposables à sa théorie et au test d'intelligence qu'il propose. Quelles sont les objections encore recevables ? Les réfutations de Turing sont-elles encore valides ? Pouvez-vous envisager des objections inédites à la lumière des nouveaux développements réalisés depuis la rédaction de cet article ? Turing y prédisait qu'un ordinateur de l'an 2000 aurait 30 % de chances de réussir à son test si celui-ci était limité à cinq minutes et administré par un examinateur non qualifié. D'après vous, qu'elles sont les chances actuelles d'un ordinateur ? Et dans cinquante ans ?

1.3 Chaque année, on décerne le prix Lœbner au programme le plus à même de réussir à une version du test de Turing. Recherchez des informations sur le dernier lauréat de ce prix. Quelles techniques a-t-il utilisées ? En quoi a-t-il fait avancer le domaine de l'IA ?

1.4 Est-ce que les actions réflexes (comme ôter sa main d'un poêle brûlant) sont rationnelles ? S'agit-il d'actions intelligentes ?

1.5 Il existe des classes de problèmes dites « impraticables » pour les ordinateurs et d'autres dont on peut démontrer l'indécidabilité. Cela signifie-t-il que l'IA est impossible ?

1.6 Supposez qu'on améliore le programme ANALOGY d'Evans de façon à ce qu'il puisse atteindre un score de 200 à un test de QI standard. Le programme serait-il plus intelligent qu'un humain ? Expliquez.

1.7 La structure neuronale de l'*Aplysia*, une limace de mer, a été largement étudiée (en premier lieu par Eric Kandel, lauréat du prix Nobel) parce qu'elle n'a que 20 000 neurones, la plupart assez gros et facilement manipulables. Si on suppose que le temps de cycle pour un neurone d'aplysie est à peu près le même que pour un neurone humain, comparez la puissance de calcul de l'aplysie en termes de transactions mémoire par rapport au superordinateur de la figure 1.3.

1.8 En quoi l'introspection (à savoir l'analyse de nos pensées intimes) peut-elle se montrer inexacte ? Puis-je me tromper sur ce que je pense ? Développez votre réponse.

1.9 Dans quelle mesure les systèmes informatiques suivants sont-ils des exemples d'intelligence artificielle :

- Lecteur de code à barres du supermarché.
- Moteurs de recherche sur Internet.
- Téléphone avec menu par reconnaissance vocale.
- Algorithme de routage de paquets pour Internet qui s'adapte dynamiquement à l'état du réseau.

1.10 Parmi les modèles informatiques qui ont été proposés pour les activités cognitives, beaucoup reposent sur des opérations mathématiques assez complexes, comme la convolution d'une image avec une gaussienne, ou la recherche du minimum d'une fonction d'entropie. La plupart des humains (et certainement la totalité des animaux) n'apprennent jamais ce genre de mathématiques, pratiquement personne ne les apprend avant le lycée et pratiquement personne ne calcule de tête la fonction de convolution avec une gaussienne. Quelle signification peut-on accorder au fait de dire que l'« appareil visuel » procède avec ce genre de mathématiques, alors que la personne elle-même n'y comprend rien ?

1.11 Certains auteurs ont affirmé que les facultés perceptuelles et motrices constituent les parties les plus importantes de l'intelligence et que les capacités de « haut niveau » sont nécessairement parasites (il ne s'agirait que de simples extensions des facultés sous-jacentes). Il est vrai que l'essentiel de l'évolution et la plus grande partie du cerveau sont consacrés aux facultés perceptuelles et motrices, alors que l'IA a trouvé des tâches comme le jeu ou l'inférence logique plus faciles que la perception et l'action dans le monde réel. Pensez-vous que l'intérêt de l'IA traditionnelle pour les capacités cognitives de haut niveau manque de pertinence ?

1.12 Pourquoi l'évolution tendrait-elle à produire des systèmes agissant de manière rationnelle ? Quels sont les buts pour lesquels ces systèmes ont été conçus ?

1.13 Est-ce que l'IA est une science, ou bien de l'ingénierie ? Ou bien ni l'une, ni l'autre, ou les deux à la fois ? Argumentez.

1.14 « Les ordinateurs ne peuvent certainement pas être intelligents ; ils ne peuvent faire que ce que leurs programmeurs leur disent de faire. » La dernière proposition est-elle vraie, et entraîne-t-elle logiquement la première ?

1.15 « Les animaux ne peuvent certainement pas être intelligents ; ils ne peuvent faire que ce que leur commandent leurs gènes. » La dernière proposition est-elle vraie, et entraîne-t-elle logiquement la première ?

1.16 « Les animaux, les humains et les ordinateurs ne peuvent certainement pas être intelligents ; ils ne peuvent faire que ce que les lois de la physique commandent à leurs atomes. » La dernière proposition est-elle vraie, et entraîne-t-elle logiquement la première ?

1.17 Étudiez la littérature sur l'IA afin de déterminer si les ordinateurs actuels peuvent mener à bien les opérations suivantes :

a. jouer correctement au ping-pong ;

b. conduire dans le centre du Caire ;

c. conduire dans Victorville, Californie ;

d. acheter les provisions d'une semaine au supermarché ;

e. acheter les provisions d'une semaine sur le Web ;

f. jouer au bridge à un niveau professionnel ;

g. découvrir et démontrer de nouveaux théorèmes mathématiques ;

h. écrire une histoire intentionnellement drôle ;

i. donner des avis pertinents dans un domaine juridique spécialisé ;

j. traduire en temps réel de l'anglais parlé en suédois ;

k. réaliser une opération chirurgicale complexe.

Pour les tâches actuellement impossibles, essayez de déterminer les difficultés qu'elles présentent et de prédire quand elles pourront être surmontées (si tant est qu'elles puissent l'être).

1.18 Plusieurs sous-domaines de l'IA ont ouvert des compétitions en définissant des tâches standard et en invitant les chercheurs à montrer leur savoir-faire. On peut citer le *DARPA Grand Challenge* pour les voitures sans pilote, la compétition internationale de planificateurs (*International Planning Competition*), la ligue de football robotique Robocup, la manifestation TREC en recherche d'information ainsi que des concours en traduction automatique et en reconnaissance de la parole. Étudiez cinq de ces événements et décrivez les progrès accomplis au fil des années. Dans quelle mesure ces compétitions ont-ils fait avancer l'état de l'art en IA ? Dans quelle mesure handicapent-ils le domaine en détournant des ressources de nouvelles idées ?