

Intelligence artificielle : entre science et marché

Quelques éléments sociohistoriques pour mieux comprendre une étrange expérimentation scientifique (1956-1990)

Par Jean-Sébastien VAYRE

Maître de conférences à l'Université Côte d'Azur

D'un côté, l'intelligence artificielle est à la mode : les principaux acteurs politiques et économiques engagés dans le développement des technologies numériques la promeuvent avec ferveur. D'un autre côté, l'intelligence artificielle fait débat : certains vont même jusqu'à proclamer qu'elle n'existe pas. L'imbrroglio est important. L'intelligence artificielle est-elle un mirage ? Dans cet article, nous verrons que non. Nous montrerons plus exactement comment s'est construite cette confusion qui caractérise l'intelligence artificielle, dont l'existence est bel et bien réelle. Pour ce faire, nous soutiendrons que l'intelligence artificielle est une discipline scientifique qui, dès le départ, a été couplée à une pratique économique produisant un déséquilibre entre recherche fondamentale et recherche appliquée. Nous ajouterons alors que c'est précisément ce déséquilibre qui est à l'origine du flou autour de cette discipline scientifique, et, plus largement, de l'instabilité caractérisant son évolution.

Introduction

À l'heure actuelle, les acteurs économiques et politiques se disputent beaucoup autour de la notion d'intelligence artificielle : ne faudrait-il pas plutôt parler d'intelligence augmentée, d'intelligence biologique, ou encore, de télé-intelligence (*remote intelligence*) ? Certains spécialistes vont même plus loin en soutenant que le mieux serait de ne plus parler d'intelligence artificielle... Sur le plan scientifique, ce type de discussion est intéressant, notamment parce qu'il a accompagné la naissance même de l'intelligence artificielle. Rappelons par exemple que Herbert A. Simon et Allen Newell n'affectionnaient pas vraiment l'expression « intelligence artificielle », proposée par John McCarthy, et préféraient celle de « système complexe de traitement de l'information » (Newell et Simon, 1956 ; Norberg, 2019 [1989]). Mais si les acteurs économiques et politiques se disputent avec autant de ferveur autour du concept d'intelligence artificielle, ce n'est pas vraiment pour des raisons scientifiques : c'est surtout parce que les enjeux du débat sont lourds de conséquences du point de vue des affaires ! Par exemple, il est assez facile de saisir pourquoi Luc Julia (2019), qui dirige le centre de recherche de Samsung, souhaite remplacer l'expression « intelligence artificielle » par « objet intelligent » : ses intérêts professionnels sont clairs puisque la stratégie d'innovation de la société qui l'emploie est entièrement tournée vers l'Internet des objets (IoT)⁽¹⁾. Aussi

paradoxal que cela puisse paraître, proclamer que « l'intelligence artificielle n'existe pas » (Julia, 2019) ne recouvre donc pas, à tout le moins dans ce cas, une simple volonté de démythification : c'est aussi et surtout une stratégie commerciale. C'est pourquoi, afin de démêler le sac de nœud que compose l'intelligence artificielle, nous proposons de reprendre les choses depuis le départ : commençons donc par nous poser les questions qui suivent. Qu'est-ce que l'intelligence artificielle ? Est-ce une science ? Est-ce un objet de consommation ? Est-ce une nouvelle forme d'intelligence autonome susceptible de dépasser celle des humains ? De notre point de vue, il est important de chercher à résoudre ce problème de définition, dans la mesure où le caractère vertueux du développement de l'intelligence artificielle dépend de la manière dont la société le comprend, c'est-à-dire du sens que nous lui donnons collectivement.

Afin de répondre aux questions qui précèdent, nous avons réalisé une enquête sociohistorique à l'aide d'ouvrages, de rapports, d'articles, de communications et de vidéos réalisés par deux grands types d'acteurs : ceux qui se sont historiquement intéressés à l'intelligence artificielle, et ceux qui ont directement participé à faire naître et exister cette discipline scientifique. Ce corpus de documents a été recueilli dans le cadre d'un travail de doctorat (Vayre, 2016), et a été enrichi au cours des recherches que nous avons effectuées ces quatre dernières années sur l'histoire de l'intelligence artificielle. Les Tableaux 1 et 2 (page 56) exposent le matériau mobilisé dans notre thèse qui est donc à

⁽¹⁾ <https://www.strategies.fr/actualites/marques/4027180W/-l-intelligence-artificielle-n-existe-pas-luc-julia.html>

l'origine des travaux présentés dans cet article. Comme nous venons de le signaler, ces travaux ont toutefois été réalisés à l'aide d'études complémentaires, et sont par conséquent fondés sur plusieurs sources qui ne sont pas mentionnées dans les Tableaux 1 et 2. Ces documents sont en revanche systématiquement référencés dans le corps de ce texte et dans sa bibliographie. En outre, notre matériau d'enquête a, dans son ensemble, été recueilli selon une méthodologie que nous pourrions qualifier d'abductive (Bruscaglioni, 2016), dans la mesure où nous avons recherché et mobilisé les documents qui le composent pour confirmer ou infirmer des hypothèses découvertes au fil de nos lectures. Autrement dit et *a contrario*, par exemple, des travaux de Dominique Cardon et de ses collègues (2018), qui s'inscrivent dans le développement de ce que nous pourrions appeler, en référence à l'École des Annales française (Burguière, 1979), une histoire quantitative de l'intelligence artificielle⁽²⁾, notre méthode de travail se situe plutôt dans la continuité du paradigme indiciaire proposé par Carlo Ginzburg (1980). Rappelons en effet que, pour cet auteur, l'histoire quantitative, si elle a le mérite de mettre en lumière les grandes structures qui font exister la dynamique d'un phénomène sur une longue période, a également tendance à l'inscrire au sein de catégories de pensée parfois trop globalisantes. Carlo Ginzburg (1980) soutient alors que les effets négatifs de cette tendance peuvent être limités en faisant appel à cette disposition d'esprit qui est, selon lui, à l'origine de l'histoire intellectuelle, et qui consiste à reconstruire une réalité invisible en interprétant les traces du passé perceptibles dans le présent. Pour reprendre les termes de Denis Thouard (2007), cette manière d'« inférer à partir des effets » constitue, à tout le moins pour Carlo Ginzburg (1980), un modèle de recherche et de réflexion particulièrement utile pour les sciences humaines et sociales. Nous avons donc cherché à nous approprier ce modèle durant l'ensemble des étapes de recherche, de recueil, de lecture et d'analyse des documents de notre matériau. En somme, les résultats de l'enquête que nous présentons dans cet article ont émergé d'un travail de sélection de documents et d'informations traduisant des morceaux de réalité empirique, que nous avons progressivement réassemblés par le biais d'effets de connaissances et d'intelligibilité caractéristique du « raisonnement sociologique » (Passeron, 1991).

Nous souhaitons maintenant préciser que, même si nous faisons quelquefois référence, dans la suite du texte, aux évolutions connues par l'intelligence artificielle ces vingt dernières années, nous nous intéressons principalement à la période qui va du milieu des années 1950 au début des années 1990, étant donné qu'elle correspond aux premiers cycles de succès et d'échecs de l'intelligence artificielle (Cardon, Cointet & Mazières, 2018). La suite de cet article expose nos résultats en deux grandes sections, qui nous conduiront à distinguer la discipline scientifique et

la pratique économique que constitue l'intelligence artificielle. Notons toutefois que, dans le sens des travaux de Bruno Latour (1987), nous avons conscience que cette distinction comporte un caractère abstrait : sur le plan purement empirique, les logiques scientifique et économique travaillant l'histoire de l'intelligence artificielle sont entremêlées par les activités des chercheurs qui la font exister. Mais c'est précisément pour démêler ces deux logiques et rendre ainsi plus aisément compréhensibles les formes de leur intrication et les enjeux y étant associés que nous proposons de structurer cet article autour de cette distinction. Aussi, nous soutenons dans la première partie de ce texte que l'intelligence artificielle est d'abord une discipline scientifique où il existe une importante concurrence, et qui vise à réaliser, de manière générale, un programme de recherche fortement expérimentale. Nous verrons alors que, si elles peuvent être considérées comme des réussites sur le plan fondamental, les recherches en intelligence artificielle constituent, à tout le moins au départ, des échecs cuisants sur le plan appliqué. Puis, nous montrons, dans la seconde partie, que l'intelligence artificielle est aussi une pratique économique où les promesses vont bon train. Nous présentons ensuite comment cette pratique joue un rôle social extrêmement important pour comprendre et expliquer les premiers cycles de succès et d'échecs caractérisant l'intelligence artificielle. En conclusion, nous dégageons quelques éléments de réflexion permettant de mieux comprendre comment s'intriquent la discipline scientifique et la pratique économique composées par l'intelligence artificielle, en soulignant comment, sur le plan sociohistorique, cette compréhension nous éclaire sur le succès actuel que connaît cette discipline.

L'intelligence artificielle comme discipline scientifique

La conférence que John McCarthy, Marvin L. Minsky, Nathaniel Rochester et Claude E. Shannon ont organisée au collège de Dartmouth en 1956 est le fondement institutionnel de l'intelligence artificielle. Dans leur proposition rédigée pour préparer cet événement, ces quatre auteurs définissent cette science de la façon suivante :

- « [L'intelligence artificielle] doit se fonder sur l'hypothèse que chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut en principe être décrit avec une telle précision qu'une machine peut être fabriquée pour le simuler. On tentera de trouver comment faire en sorte que les machines utilisent le langage, produisent des abstractions et des concepts, résolvent des types de problèmes désormais réservés aux humains et s'améliorent. Nous pensons qu'une avancée significative peut être réalisée dans un ou plusieurs de ces problèmes si un groupe de scientifiques soigneusement sélectionnés y travaille ensemble pendant un été » (traduit par nous : McCarthy *et al.*, 2019 [1955]).

⁽²⁾ Même si les auteurs mobilisent quelques entretiens pour retracer l'histoire de l'intelligence artificielle, ils s'appuient en effet principalement sur l'analyse statistique d'un matériau de plus de 27 000 articles recueillis en 2018 sur *Web of Science*.

Types d'approche	Références
Intelligence artificielle de type symbolique	(Bickhard et Terveen, 1995), (Bonissone et Johnson, 1984), (Fodor, 1975), (Fodor, 1983), (Fodor et Pylyshyn, 1988), (Forgy, 1981), (Gaschnig, 1980), (Ince, 1992), (Laird, Newell & Rosenbloom, 1987), (Lenat, 1977), (Lenat, 1983), (Lindsay, Buchanan, Feigenbaum & Lederberg, 1993), (McCarty, 1977), (Memmi, 1990), (Minsky et Papert, 1969), (Neumann, 1958), (Newell, 1980), (Newell et Simon, 1972), (Papert, 1988), (Samuel, 1959), (Simon, 1991 [1969]), (Simon, 1992), (Tristan et Abdallah, 2009), (Turing, 1950), (Winston, 1970)
Intelligence artificielle de type connexionniste	(Ackley, Hinton & Sejnowski, 1985c), (Bechtel et Abrahamsen, 1993), (Bickhard et Terveen, 1995), (Changeux, 1983), (Fodor et Pylyshyn, 1988), (Hayek, 1952), (Hebb, 1949), (Hopfield, 1982), (Lai, 2015), (LeCun et Bengio, 1995), (LeCun <i>et al.</i> 1989), (McCulloch et Pitts, 1943), (Memmi, 1990), (Minsky et Papert, 1969), (Noduls, 2015) ; (Numenta, 2011), (Numenta, 2014), (Numenta, 2015), (Papert, 1988), (Rosenblatt, 1958), (Rumelhart, Hinton & Williams, 1986), (Smith, 1999), (Smolensky, 1988), (Widrow et Hoff, 1960)
Intelligence artificielle de type hybride	(Bonasso, Firby, Gat, Kortenkamp, Miller & Slack, 1997), (Cassimatis, 2005), (DePristo et Zubek, 2001), (Hawes <i>et al.</i> 2007), (Kubera, Mathieu & Picault, 2011), (Langley et Choi, 2006), (Müller et Pischel, 1993), (Reynaud, 2014), (Schmidt, 2005), (Silver <i>et al.</i> 2016), (Smolensky, 1987), (Smolensky, Legendre & Miyata, 1992)

Tableau 1 : Documents produits par des chercheurs ou des organisations directement engagés dans le développement de l'intelligence artificielle⁽³⁾.

Types de matériau	Références
Documents écrits	(Blanc, Charron & Freyssenet, 1989), (Boise, 2007), (Copeland et Proudfoot, 2015), (Dupuy, 1994), (Hodges, 2014 [1983]), (Pélissier et Tête, 1995), (Varela, 1988)
Documents vidéo	(Dammbeck, 2003), (Folgoas, 1976), (Guirardoni, 1981), (Karlin, 1971), (Lallier, 1963), (Moreuil, 1972), (Royer, 1961a), (Royer, 1961b)

Tableau 2 : Documents produits par des chercheurs et des journalistes qui se sont historiquement intéressés à l'intelligence artificielle.

Rappelons toutefois que les origines de l'intelligence artificielle s'inscrivent plus largement au sein de l'histoire de l'informatique et de la cybernétique. Par exemple, les travaux de Blaise Pascal, de Gottfried W. Leibniz, de Charles Babbage, de Augusta Ada King, de George Boole, de Friedrich L. G. Frege, de Kurt Gödel et, bien sûr, de Alan M. Turing, de John von Neumann, ou encore ceux de Norbert Wiener, de Warren McCulloch et de Walter H. Pitts ont joué un rôle fondamental dans l'émergence de cette science (Crevier, 1997 [1993] ; Pratt, 1995 [1987] ; Rose, 1986 [1984]). En outre, il est important de rappeler que, dès le début, l'appellation « intelligence artificielle » pour désigner cette nouvelle science ne fait pas l'unanimité. Elle plaît avant tout à John McCarthy, qui finit progressivement par persuader ses collègues de l'accepter.

Comme nous l'avons mentionné, Herbert A. Simon et Allen Newell préfèrent parler de « système complexe de traitement de l'information » (*complex information processing system* ; Newell et Simon, 1956 ; Norberg, 2019 [1989]).

Des styles de recherche différents

À l'instar des autres sciences, l'intelligence artificielle ne recouvre pas une communauté d'entente parfaite : elle ne constitue pas un ensemble harmonieux de représentations que chaque participant partage collectivement. Pour reprendre les termes de Pierre Bourdieu (1976), l'intelligence artificielle constitue un champ scientifique dont la stabilité dépend des jeux de force qui l'animent, et à l'intérieur desquels se font et se défont diverses formes de domination. Ceci est d'ailleurs d'autant plus vrai que l'intelligence artificielle est fortement interdisciplinaire : en fonction de leurs appétences, les chercheurs qui la font peuvent croiser des sciences aussi différentes que la biologie, la

⁽³⁾ Le détail de l'ensemble des références exposées dans les Tableaux 1 et 2 est consultable dans la bibliographie de notre thèse (Vayre, 2016).

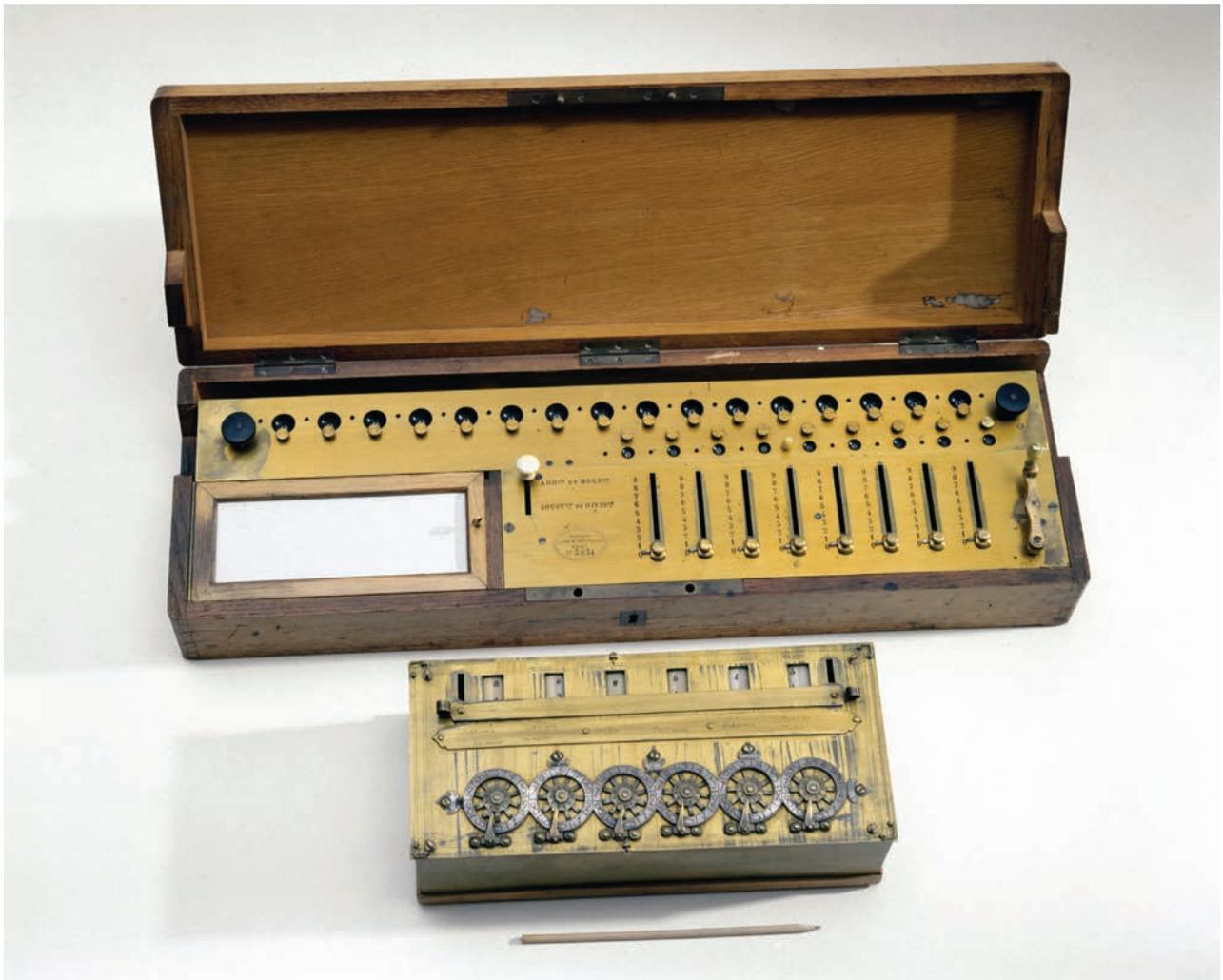


Photo © SSPLUJIG / BRIDGEMAN IMAGES

Machine à calcul de Pascal (1642) et arithmomètre de Colmar (vers 1850). Londres Science Museum.

« Les origines de l'intelligence artificielle s'inscrivent plus largement au sein de l'histoire de l'informatique et de la cybernétique. Les travaux de Blaise Pascal, de Gottfried W. Leibniz, de Augusta Ada King, de Kurt Gödel et, bien sûr, d'Alan Turing ou de John von Neumann, [...] ont joué un rôle fondamental dans l'émergence de cette science. »

psychologie, l'anthropologie, la logique, la philosophie, la linguistique, les mathématiques, l'électronique et l'informatique. Il n'en reste pas moins que l'intelligence artificielle s'est constituée autour d'un objectif commun. L'ensemble des chercheurs engagés dans cette science a contribué à tester l'hypothèse selon laquelle une machine peut réaliser des comportements que les hommes considèrent généralement comme intelligents. Et, dès le départ, les manières de participer à cette expérimentation font l'objet d'âpres discussions.

Lors de la conférence de Dartmouth, les chercheurs qui se démarquent le plus de la communauté réunie sont Herbert A. Simon et Allen Newell ; et cela, pour plusieurs raisons. Premièrement, en 1956, Herbert A. Simon et Allen Newell sont les seuls à disposer d'un programme informatique capable de réaliser artificiellement une des dimensions de l'intelligence que les universitaires considèrent souvent comme des plus nobles, c'est-à-dire celle impliquée dans la résolution de problèmes mathématiques complexes. « Logic Theory machine » (LT ; Newell et Simon, 1956) est en effet capable de prouver plus de la moitié des théorèmes des *Principia Mathematica* de

Alfred N. Whitehead et de Bertrand A. W. Russell. Deuxièmement, LT est conçu à partir de connaissances issues du champ des sciences humaines et sociales puisqu'il encapsule une partie des idées à la base de la théorie de la rationalité limitée (Simon, 1945)⁽⁴⁾. Or la plupart des chercheurs présents à Dartmouth considèrent qu'il n'est pas utile d'étudier les processus cognitifs humains pour concevoir un programme d'intelligence artificielle. C'est le cas de John McCarthy et de Marvin L. Minsky, qui, à la fin des années 1950, partagent l'idée que l'intelligence artificielle doit se concentrer sur l'exploration de la logique formelle. Mais ce point de vue est tout aussi controversé que celui d'Herbert A. Simon et d'Allen Newell. Par exemple, Herbert Gelernter et Nathaniel Rochester (1958), ou encore Oliver G. Selfridge (1959) appréhendent l'intelligence artificielle selon d'autres perspectives. Pour eux, cette science ne doit pas se constituer en référence à la cognition humaine, ou encore à la logique formelle. Elle doit se fonder sur les seules capacités de traitement informatique des machines. Cette approche

⁽⁴⁾ Rappelons que cette théorie a permis à Herbert A. Simon d'obtenir le prix Nobel d'économie de 1978.

leur permet alors de développer leurs premiers programmes d'intelligence artificielle : dans le cas d'Herbert Gelernter et de Nathaniel Rochester, il s'agit du « Geometry Theorem Prover » (GTP ; Gelernter et Rochester, 1958), et dans celui de Oliver G. Selfridge du fameux « Pandemonium » (Selfridge, 1959).

Trois grands axes de tension

Dès sa naissance, l'intelligence artificielle est traversée par de nombreuses tensions. Avec le temps, ces tensions s'amplifient progressivement pour finir par structurer durablement le champ de l'intelligence artificielle. Entre les années 1960 et les années 1990, il existe au moins trois grands axes de tension qui ont joué un rôle crucial dans la dynamique de cette science.

Le premier renvoie aux désaccords cordiaux qui émergent rapidement entre John McCarthy et Marvin L. Minsky concernant la manière dont il est possible de définir le problème de l'intelligence artificielle : rappelons en effet que pour John McCarthy, la question de fond posée par cette nouvelle discipline est principalement d'ordre logique. Or, Marvin L. Minsky ne partage pas cet avis (Norberg, 2019 [1989]). Dès 1960, un premier axe de tension naît ainsi entre ces deux chercheurs, qui se sont, à partir de là, constamment engagés sur des chemins de recherches différents. De telle sorte qu'en 1962, John McCarthy décide de partir du laboratoire d'intelligence artificielle du Massachusetts Institute of Technology (MIT AI Lab) pour aller diriger son propre laboratoire à l'Université de Stanford (le *Stanford Artificial Intelligence Lab* ou SAIL). Précisons alors qu'à ce moment-là ses travaux sur la logique ont un retentissement important dans la communauté des spécialistes de l'intelligence artificielle. Le langage *list processing* (LISP) qu'il conçoit en 1958 permet, par exemple, à Douglas Lenat d'élaborer ses programmes « Automated Mathematician » (AM ; Lenat, 1977) et EURISKO (Lenat, 1983). De même, la structure dite « Advice Taker » de type « SI, ALORS » que John McCarthy imagine en 1959 joue un rôle important dans le développement des systèmes experts, tout comme les travaux qu'il conduit quelques années plus tard sur la « circonscription »⁽⁵⁾, fluidifiant le traitement d'information qu'effectuent ces systèmes (Crevier, 1997 [1993]). Il n'en reste pas moins que, à la suite du départ de John McCarthy, Marvin L. Minsky se retrouve seul à la tête du MIT AI Lab ; une position plutôt confortable puisqu'il bénéficie de cette façon d'importants investissements pour faire face à son nouveau concurrent : afin de financer le projet MAC (*Machine-Aided Cognition and Multiple Access Computer*), la Defense Advanced Research Projects Agency (DARPA) verse trois millions de dollars pendant plusieurs années au MIT (Flamm, 1987 ; Norberg, 2019 [1989]). L'ancien *alter ego* de John McCarthy dispose ainsi de ressources non négligeables pour fonder le courant antilogique de l'intelligence artificielle. Partant, compte tenu des

moyens économiques, techniques et humains dont dispose le MIT AI Lab, ce courant ne tarde pas à connaître un succès important auprès de la communauté des pairs. De nombreux jeunes et brillants chercheurs sont de cette manière conduits à travailler avec Marvin L. Minsky. C'est le cas de James R. Slagle, de Joel Moses, de Patrick Winston et de Seymour A. Papert qui élaborent respectivement les programmes « Symbolic Automatic INTEgrator » (SAINT ; Slagle, 1961), « Symbolic Integration » (SIN ; Moses, 1967), le programme d'apprentissage du concept d'arche (Winston, 1970) et le langage informatique LOGO (Papert, 1971).

En somme, comme le montre la Figure 1 (page suivante), la séparation du couple McCarthy/Minsky donne naissance à deux grandes manières de faire de l'intelligence artificielle. Pour Roger C. Schank, il existe, au moins jusqu'au début des années 1990, deux styles différents de recherche en intelligence artificielle : le style des « tirés à quatre épingles », qui renvoie plutôt au courant logique développé par John McCarthy, et le style des « as de pique », plutôt associé au courant antilogique impulsé par Marvin L. Minsky :

- « Pour Schank, les tirés à quatre épingles portent des vêtements bien propres et travaillent sur des phénomènes de surface, comme la logique ou la syntaxe, qu'ils sont à même de comprendre et de ranger dans de petites cases douillettes. Les as de pique s'habillent au petit bonheur et adorent se débattre avec des problèmes tortueux comme la sémantique » (Crevier, 1997 [1993], p. 201).

À ce premier axe de tension peut alors être ajouté un deuxième qui permet de distinguer le courant de recherche initié par Roger C. Schank de celui de ses collègues Herbert A. Simon et Allen Newell. Dans les années 1960, ces derniers exercent déjà leurs fonctions d'enseignement et de recherche à l'Université de Carnegie Mellon, dotée d'une excellente réputation auprès des universitaires, mais aussi des industriels (Norberg, 2019 [1989]). Herbert A. Simon et Allen Newell sont de ce fait rapidement reconnus pour leurs travaux sur la résolution de problèmes. À la suite de l'effet de surprise qu'à produit LT, Herbert A. Simon et Allen Newell continuent d'étudier et de s'inspirer de la cognition humaine pour concevoir de nouveaux programmes informatiques. Ils sont par là même conduits à développer leur fameux « General Problem Solvers » (GPS ; Newell, Shaw et Simon, 1959) qui est à l'origine de la conception des systèmes experts les plus reconnus. Rappelons par exemple que Edward A. Feigenbaum et Bruce G. Buchanan s'inspirent directement de GPS pour développer les systèmes experts que sont DENDRAL (Buchanan et Feigenbaum, 1978) et MYCIN (Buchanan et Shortliffe, 1984), tout comme leurs étudiants Randall Davis et John P. McDermott qui élaborent respectivement les programmes TEIRESIAS (Davis, 1978) et « eXpert CONFIGurer » (XCON ; McDermott, 1982). Or, les travaux que Roger C. Schank (1978) effectuent à l'Université de Yale ne portent pas sur la résolution de problèmes d'experts. Si Roger C. Schank partage avec Herbert A. Simon et Allen Newell l'idée que

⁽⁵⁾ Cette dénomination renvoie à un procédé informatique dont la fonction est d'isoler ou de minimiser les obstacles que peuvent rencontrer les moteurs d'inférence logique, pour naviguer à l'intérieur d'un système à base de connaissance.

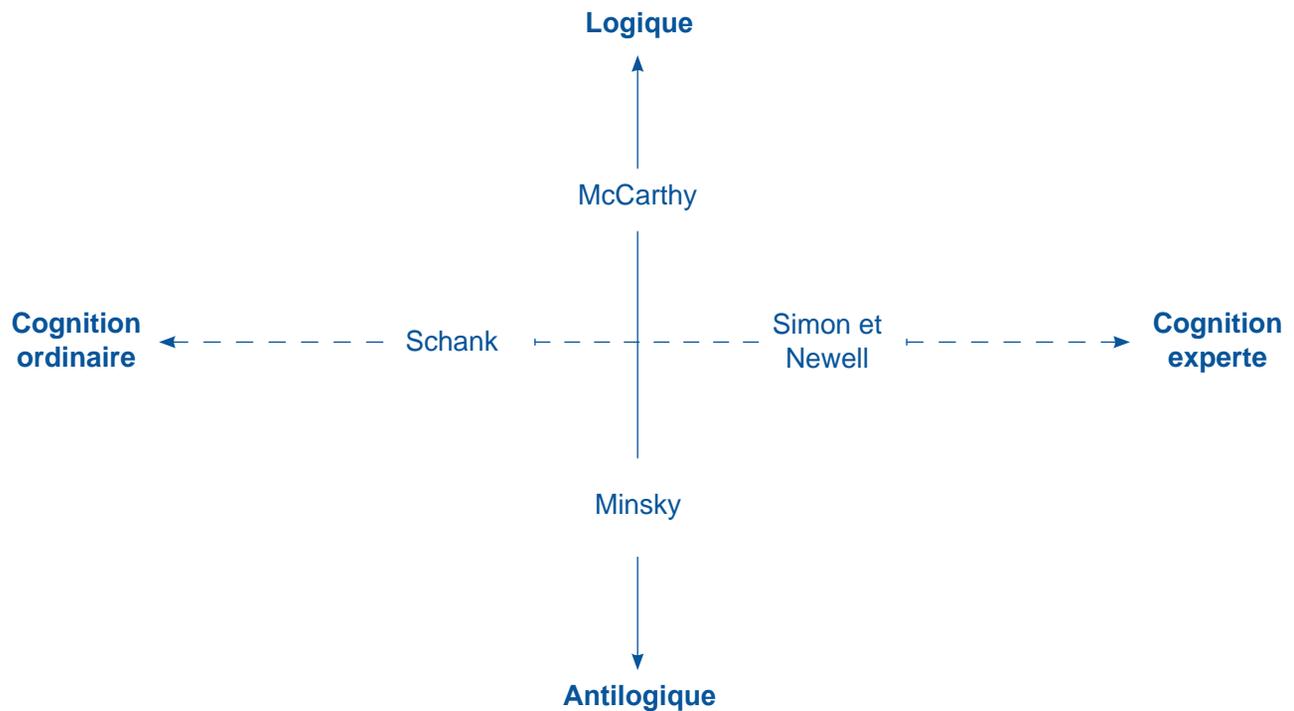


Figure 1. Les quatre principaux courants de recherche en intelligence artificielle symbolique.

l'intelligence artificielle doit se fonder sur l'étude de la cognition humaine, il s'intéresse moins à l'intelligence dite de haut niveau qu'à celle que tout un chacun met en œuvre pour agir au quotidien. Autrement dit, Roger C. Schank n'étudie pas la cognition experte. Il s'intéresse à la cognition ordinaire. Son objectif n'est pas de développer des programmes informatiques pour permettre aux machines d'aider ou de remplacer des personnes qui sont hautement qualifiées. Il cherche plutôt à comprendre comment ces machines peuvent s'adapter au quotidien des hommes pour les accompagner et les aider à mieux vivre en société.

Dès lors, il convient de préciser que, comme le montrent les travaux de Dominique Cardon et de ses collègues (2018), il existe également un troisième axe de tension, qui a joué un rôle fondamental dans l'histoire de l'intelligence artificielle. Ce troisième axe oppose les versants dits symbolique et connexionniste de cette discipline. Car les deux axes de tension dont nous venons de parler renvoient au versant symbolique de l'intelligence artificielle : les chercheurs précités partagent, dans l'ensemble, l'idée que l'intelligence constitue un système de computation de représentation symbolique plutôt d'ordre délibératif (Fodor et Pylyshyn, 1988). Et c'est précisément pourquoi, à ce moment-là et à la suite d'Herbert A. Simon et d'Allen Newell, de nombreux chercheurs portent un intérêt particulier à la figure de l'expert. Or, si cette approche domine le champ de l'intelligence artificielle entre les années 1960 et les années 1990, un basculement s'opère depuis les années 2000 (Cardon, Cointet & Mazières, 2018 ; Vayre, 2016). Le versant connexionniste de l'intelligence artificielle, que Frank Rosenblatt (1958), Wilfrid K. Taylor (1956), Bernard Widrow et Marcian E. Hoff (1960) ont beaucoup de mal à défendre

face aux critiques de Marvin L. Minsky et de Seymour A. Papert (1969)⁽⁶⁾, devient aujourd'hui dominant, comme l'indique le succès du *deep learning*, mais aussi celui des théories neurologiques pour expliquer les comportements individuels et collectifs (Changeux, 1983). Et, de ce point de vue, l'intelligence compose un gigantesque réseau d'activations et d'inhibitions d'unités de calculs élémentaires, qui est plutôt d'ordre adaptatif. Partant, l'attention collective des chercheurs en intelligence artificielle est moins dirigée vers le raisonnement de l'expert que vers ce que nous pourrions qualifier, de façon schématique, d'« apprentissage de l'enfant » (Vayre, 2016). Aussi, cet autre versant de l'intelligence artificielle, qui est fondé sur le caractère biologique de la cognition et mathématique du traitement de l'information, recouvre, lui aussi, différents styles de recherche. Afin de ne pas disperser nos propos, nous préférons toutefois laisser dans l'ombre, à tout le moins dans ces travaux, cet autre versant de l'intelligence artificielle.

Des succès et des échecs : le cas du laboratoire BAIR

Dès le départ, les réussites réalisées dans le domaine de l'intelligence artificielle sont nombreuses : les programmes fabriqués par les chercheurs issus de cette discipline effectuent ce pourquoi ils sont conçus. Par exemple, le programme SHRDLU, que conçoit Terry Winograd (1972), est capable de s'interroger sur ses motivations et de disposer ainsi de quelque chose qui pourrait s'apparenter à une certaine « conscience

⁽⁶⁾ Précisons toutefois que, malgré ces critiques, ces deux auteurs portent un intérêt certain au connexionnisme. Il n'est en ce sens peut-être pas inutile de rappeler que Marvin L. Minsky (1954) a fait sa thèse sur les réseaux de neurones.

de soi » ; deux propriétés qui lui permettent – et c’est une des premières fois dans le domaine de l’intelligence artificielle – de réaliser un dialogue réellement constructif avec l’utilisateur. Pour illustration, voici le type d’échanges qu’un utilisateur peut entretenir avec SHRDLU⁽⁷⁾ :

- « 23. Avez-vous touché une pyramide avant de mettre la [pyramide] verte sur le petit cube ? OUI, LA VERTE [...]. 24. Quand l’avez-vous ramassée ? PENDANT QUE J’EMPILAIS LE CUBE ROUGE, UN GRAND BLOC ROUGE ET UN GRAND CUBE VERT [...]. 25. Pourquoi ? [...] POUR M’EN DÉBARRASSER [...]. 26. Pourquoi avez-vous fait cela ? POUR LIBÉRER LE CUBE ROUGE [...]. 29. Pourquoi avez-vous fait cela ? PARCE QUE VOUS ME L’AVEZ DEMANDÉ » (traduit par nous ; Winograd, 1972, p. 13).

Aussi impressionnant que cela puisse paraître sur le plan fondamental, le problème est que ce dialogue ne peut être véritablement qualifié de constructif qu’à l’intérieur des frontières du micro-monde dans lequel évolue SHRDLU, c’est-à-dire un environnement virtuel fait de quelques cubes, cônes et sphères ou, si l’on préfère, un « monde jouet » (Cardon, Cointet & Mazières, 2018). Nous l’avons déjà dit, l’intelligence artificielle est une discipline expérimentale. Ainsi, chaque réussite, aussi importante soit-elle, est rapidement rattrapée par un échec : les progrès effectués sont systématiquement des manières de rappeler à ceux qui les réalisent combien les objectifs qu’ils visent sont loin d’être atteints. Afin de bien comprendre ce dernier point, nous proposons maintenant de nous centrer sur le cas du laboratoire de recherche en intelligence artificielle de l’Université de Berkeley (Berkeley Artificial Intelligence Research ou BAIR). Ce dernier constitue en effet un cas aussi révélateur qu’intéressant pour saisir les limites et les avantages du caractère expérimental de l’intelligence artificielle.

Dès le début des années 1980, le BAIR est dirigé par Robert Wilensky, un ancien étudiant de Roger C. Schank. Dans la continuité de ce dernier, que l’on qualifie souvent d’« enfant terrible de l’intelligence artificielle », Robert Wilensky n’aime pas les formalismes, peu importe qu’ils soient d’ordre logique ou d’ordre mathématique. Contrairement à nombre de ses collègues qui voient dans les travaux de Noam Chomski (1965) la possibilité de formaliser l’intelligence humaine, il ne croit pas que le langage puisse se réduire à une syntaxe formelle. S’il partage l’idée du paradigme chomskien selon lequel le langage est au cœur de la pensée, il pense également que ce dernier compose un problème d’ordre sémantique, et non d’ordre syntaxique. En d’autres termes, pour reprendre l’expression de Roger C. Schank, Robert Wilensky est un as de pique ! Il aime les problèmes épineux et élabore une perspective de l’intelligence artificielle qui est à son image : à la fois originale et audacieuse. Car Robert Wilensky est un personnage atypique et souvent considéré comme

hérétique par une bonne partie de ses collègues (Rose, 1986 [1984]). Il aime l’originalité, et c’est probablement ce goût qui le conduit à apprécier le climat intellectuel régnant à l’Université de Berkeley : si cette dernière n’a pas vraiment de culture informatique lors de l’arrivée de Robert Wilensky, elle cultive en revanche une diversité intellectuelle qui lui plaît beaucoup. À Berkeley, il y a de tout : des anthropologues idiosyncratiques, des linguistes hérétiques, des psychologues cognitivistes... Et surtout, il y a Hubert L. Dreyfus et John R. Searle qui participent activement à stimuler les recherches de Robert Wilensky : en critiquant sans cesse les avancées de l’intelligence artificielle, ces deux philosophes contribuent en effet à dynamiser l’activité du BAIR, mais aussi à donner de l’importance à son directeur. Pour Robert Wilensky, qui s’intéresse au sens commun et au problème de l’action ordinaire, Berkeley constitue l’endroit idéal pour expérimenter de nouvelles façons de concevoir l’intelligence artificielle. Dès lors, afin de bien saisir le caractère expérimental du projet de recherche réalisé au BAIR, il nous faut maintenant exposer quelques-uns des principaux programmes conçus par Robert Wilensky et ses collègues.

Un peu à la manière du « Script Applier Mechanism » (SAM) élaboré par Roger C. Schank et Robert P. Abelson (1977), le « Plan Applier Mechanism » de Robert Wilensky (1977) manifeste une certaine capacité à comprendre les récits et les situations qui lui sont décrits. En 1980, PAM est par exemple capable de réaliser l’échange qui suit :

- « [À partir de la description suivante :] “John a besoin d’argent, il prend un revolver et va dans un bar. Il dit au propriétaire qu’il veut de l’argent ; le propriétaire donne de l’argent à John qui s’en va” [...] et de la question suivante :] “Pourquoi le propriétaire donne-t-il de l’argent à John ?” [...] PAM répondait par exemple :] “Le propriétaire avait peur que John le tue” » (Rose, 1986 [1984], p. 71).

Comme SAM, PAM fait preuve d’un certain niveau de compréhension dans le sens où il manifeste un savoir qui n’est pas explicitement présent dans l’énoncé. Mais PAM dispose de quelque chose de plus que SAM : pour comprendre une situation, PAM n’a pas besoin que son concepteur lui fournisse le scénario sous-jacent. Bien sûr, un peu à la manière du programme SHRDLU de Terry Winograd (1972), PAM n’est capable d’une telle prouesse qu’à la condition que les énoncés communiqués concernent son micro-monde. Et, comme le montre Frank Rose (1986 [1984]) à partir du cas du « Plan ANalyzer with Dynamic Organization, Revision and Application » (PANDORA), étendre ce micro-monde recouvre une myriade d’astuces informatiques, qui sont autant de façons de questionner le fonctionnement de l’intelligence humaine. Par exemple, Joe Faletti (1982), l’étudiant de Robert Wilensky qui a conçu PANDORA, éprouve beaucoup de difficulté pour faire comprendre à son programme que l’action d’aller chercher le journal dans la boîte aux lettres peut renvoyer à différents comportements en fonction du temps qu’il fait. Sur le plan cognitif, une telle capacité

⁽⁷⁾ Les communications écrites en lettres capitales correspondent aux phrases exprimées par SHRDLU.

de compréhension et d'adaptation renvoie effectivement à des problèmes de planification importants, notamment en termes d'organisation des buts et des sous-buts de l'action, des tâches qui la composent, mais aussi en termes de mémorisation des informations pertinentes, c'est-à-dire d'organisation et d'activation des connaissances utiles à l'effectuation de l'ensemble de ces tâches (Faletti, 1982). Par exemple, pour que PANDORA soit capable de mettre un imperméable, il faut qu'il reconnaisse que la pluie mouille et qu'être au sec est un état désirable, mais il faut aussi qu'il sache qu'un imperméable protège de la pluie. Et, aussi étrange que cela puisse paraître, pour Joe Faletti, ce type de problème est aussi important que difficile à résoudre sur le plan informatique. C'est d'ailleurs ici que se trouve le triple avantage des travaux de Robert Wilensky et de ses collègues. Dans leur ensemble, ces derniers mettent en évidence que :

- aussi ordinaire soit-elle, l'action recouvre des formes de résolution de problème qui, malgré leur caractère automatique, sont complexes sur le plan cognitif ;
- ces formes de résolution de problème sont indissociables des conventions sociales qui préexistent à l'action ;
- l'articulation du cognitif et du social à l'œuvre dans l'accomplissement de toute action humaine (même si elle est des plus triviales) recouvre une forme d'intelligence extrêmement difficile à identifier, à décrire, à comprendre et à formaliser.

À l'encontre des discours simplistes que l'on peut parfois entendre sur l'intelligence artificielle, le cas du laboratoire BAIR nous montre donc que l'intelligence artificielle ne se réduit pas à une communauté de chercheurs souhaitant imposer leur point de vue de logiciens et de mathématiciens en l'appliquant au domaine des sciences humaines et sociales. Pour Robert Wilensky et ses collègues, l'ordinateur est un outil d'expérimentation scientifique qui a la vertu heuristique de les aider à mettre en question et à comprendre ce qu'est l'intelligence. Mais ce point de vue n'est pas spécifique au BAIR : par exemple, comme nous l'avons vu, Herbert A. Simon, Allen Newell et leurs étudiants de Carnegie Mellon partagent cette idée. Pour nous, le cas du laboratoire BAIR comporte un autre intérêt. Celui-ci réside dans cette vision à la fois critique et originale que Robert Wilensky et ses collègues ont de l'intelligence artificielle. À la suite des travaux de Marvin L. Minsky (1974) sur les schémas (*frames*), PANDORA⁽⁸⁾ constitue une manière de représenter informatiquement une conception de l'intelligence qui raisonne avec la théorie du "*Background*" de John R. Searle (2002), c'est-à-dire celle selon laquelle le langage est un code dont la signification ne peut exister sans les conventions sociales qui permettent son expression. Ainsi, l'histoire a montré que Robert Wilensky et son équipe échouent dans leur projet de conception d'un programme informatique capable de simuler la cognition ordinaire. Peut-on pour autant en conclure que leur programme

de recherche est un échec ? Cela dépend du positionnement de l'acteur qui répond. Pour un investisseur comme la DARPA ou International Business Machine (IBM), la réponse est positive : un programme informatique qui ne fonctionne que dans le cadre d'un micro-monde inventé par un chercheur n'a aucune utilité politique ou économique. À l'inverse, pour le sociologue qui s'intéresse à l'histoire des sciences et des techniques, la réponse est clairement négative. Bien sûr, par le biais de son expérimentation, le BAIR n'a pas réussi à valider l'hypothèse selon laquelle il est possible de représenter informatiquement la cognition ordinaire. Mais, de notre point de vue, cet échec constitue en réalité une immense réussite, puisque c'est grâce à cet insuccès que l'équipe de Robert Wilensky a pu mieux expérimenter la complexité des interactions qui se jouent entre le cognitif et le culturel, la difficulté qui existe à représenter informatiquement cette complexité, mais aussi et surtout l'idée que la moindre action ordinaire, pour être correctement réalisée, interprétée et comprise, implique l'effectuation de cette complexité.

Aussi fragile soit-elle, la valeur cognitive derrière une telle conclusion est donc particulièrement élevée puisque, dans la continuité des travaux d'Herbert A. Simon et d'Allen Newell, elle donne lieu de remettre en question la manière dont les hommes conçoivent l'intelligence. Les expérimentations réalisées au BAIR permettent, par exemple, de formuler l'hypothèse selon laquelle, si Herbert A. Simon et Allen Newell ont rapidement réussi à développer des simulations satisfaisantes de la cognition experte, c'est parce que, contrairement à ce que la majorité des universitaires peuvent penser, celle-ci est probablement plus idiote qu'elle n'y paraît. Même si cela peut paraître extravagant, pour Robert Wilensky et ses collègues, il n'est effectivement pas si sûr que résoudre la moitié des *Principia Mathematica* nécessite une plus grande intelligence qu'aller chercher son courrier dans sa boîte aux lettres.

L'intelligence artificielle comme pratique économique

Nous avons vu que l'intelligence artificielle s'inscrit dans la continuité de la cybernétique et de l'histoire des idées qui la fondent. Cela veut dire que l'intelligence artificielle émerge avec le développement des premiers ordinateurs : sa naissance est concomitante à celle de l'informatique qui, rappelons-le, recouvre un domaine d'activité à la fois scientifique, technique et industriel. Il est en ce sens important de garder en tête que l'intelligence artificielle est indissociable des grandes innovations sociotechniques qui ont permis l'informatisation de la société (Mounier-Kuhn, 2010). Et, dans les années 1950, ces innovations coûtent cher ! De telle sorte que les pionniers de l'intelligence artificielle ont dû rapidement s'allier aux acteurs politiques et économiques pour financer leurs recherches. Notons qu'il n'y a jusque-là rien de très nouveau puisque les cybernéticiens ont fait de même : Alan M. Turing s'associe au gouvernement anglais pour décrypter ENIGMA et aider les alliés à combattre les nazis

⁽⁸⁾ Comme le programme PAMELA qu'a conçu Peter Norvig pour le compléter (Faletti, 1982).

(Hodges, 2015 [1983]), John von Neumann s'allie au gouvernement américain pour optimiser l'explosion de la bombe atomique et permettre aux Américains d'intimider l'Union soviétique (Hoddeson *et al.*, 1992), et, de manière plus générale, à la suite des deux guerres mondiales, l'ensemble des mathématiciens et des cybernéticiens présents aux conférences de Macy souhaitent contribuer à l'instauration d'un nouvel ordre mondial, qui doit garantir la paix entre les hommes, mais aussi leur « santé mentale », c'est-à-dire leur autonomie et leur liberté intellectuelle (Heims, 1991).

Autrement dit, dans la continuité des arguments exposés en amont, nous souhaitons maintenant insister sur le fait que, pour faire carrière dans le domaine de l'intelligence artificielle, il ne suffit pas d'être un chercheur reconnu et apprécié par ses pairs. Il faut également savoir attirer les investisseurs afin de capter le plus longtemps possible leur financement (Latour, 1987). Pour ce faire, les chercheurs en intelligence artificielle ont été amenés à se positionner sur les plans politique et économique, notamment parce que la DARPA constitue pendant longtemps la plus importante source de financement de cette science. Et c'est par le biais de cette forme particulière d'« économicisation » (Akrich, 1989) que l'intelligence artificielle a pu entrer dans les débats publics. Plusieurs controverses ont ainsi émergé et ont donné lieu à des débordements qui ont obscurci la compréhension collective de ce qu'est l'intelligence artificielle. C'est donc afin de mieux saisir comment cette obscurité a pu advenir, mais aussi quels sont ses effets sur le développement de cette science, que nous souhaitons revenir sur l'histoire des relations qui ont existé entre l'intelligence artificielle et le marché.

Des promesses pour attirer les investisseurs

L'intelligence artificielle est un champ scientifique tirailé par les trois grands axes de tension exposés précédemment (*cf.* section « Des déceptions et des peurs »). Il faut alors bien comprendre que, pour lutter au sein d'un tel champ de force, les chercheurs issus de cette discipline ont dû trouver des alliés capables de leur fournir les moyens techniques et financiers leur permettant de réaliser leurs programmes de recherche. Ils ont de cette façon été collectivement conduits à instituer des passerelles entre les mondes scientifique, politique et économique afin de pouvoir faire exister les tensions susmentionnées. Nous avons vu que, dès le départ, l'intelligence artificielle forme un environnement de travail tumultueux : si les chercheurs qui le composent savent se montrer sympathiques, notamment afin de trouver les alliés qui les aideront à réaliser au mieux leurs travaux, ils ont également conscience qu'ils vont devoir jouer des coudes pour atteindre leurs objectifs de carrière. Les grandes écoles et universités qui disposent d'un laboratoire en intelligence artificielle favorisent d'ailleurs ce climat de compétition, pour au moins deux raisons. La première est que, comme nous l'avons vu avec le cas du MIT AI Lab, ces laboratoires peuvent parfois garantir plusieurs millions de dollars par an aux écoles et aux universités qui les accueillent.

La seconde est que, compte tenu de l'intérêt porté par le complexe militaro-industriel à l'intelligence artificielle, ces mêmes écoles et universités travaillent dur pour attirer les chercheurs qui bénéficient des meilleures réputations. C'est ainsi que, dans le domaine de l'intelligence artificielle, la science et le marché finissent par s'auto-alimenter pour donner forme à un type d'« économie d'opinion » (Orléan, 2000) particulier où ce qui compte n'est plus seulement la valeur scientifique du programme de recherche conduit par le laboratoire : c'est aussi et surtout la capacité des chercheurs qui le composent à faire preuve d'autorité sur leurs pairs, à provoquer positivement l'intérêt de la société, et à susciter l'attention et la confiance des acteurs politiques et économiques.

Au sein de cet environnement de haute compétition, l'arrogance peut quelquefois faire la différence. C'est, du moins, ce que semble montrer les cas d'Herbert A. Simon et d'Allen Newell, qui, rappelons-le, ont toujours bénéficié d'une importante reconnaissance dans le domaine de l'intelligence artificielle. Les deux compères de Carnegie Mellon sont en effet connus pour leur caractère vaniteux. Par exemple, lors d'un entretien, Marvin L. Minsky confie à Daniel Crevier (1997 [1993]) qu'Herbert A. Simon et Allen Newell se sont montrés distants lors de la conférence de Dartmouth. Aux yeux des autres participants, ces deux chercheurs ont semblé aussi flattés que satisfaits d'être les seuls à présenter un programme d'intelligence artificielle. Herbert A. Simon lui-même confirmera cette description (Crevier, 1997 [1993], p. 67). Ajoutons qu'en 1997, soit quarante ans après Dartmouth, la confiance qu'Herbert A. Simon et Allen Newell ont en leurs travaux ne s'affaiblit pas, bien au contraire. Dans son ouvrage de 1991, Herbert A. Simon exprime l'idée qu'en inventant un programme informatique capable de traiter des données symboliques, Allen Newell et lui-même ont révélé comment un système composé de matière peut exprimer des propriétés de la pensée. Pour Herbert A. Simon (1991), leurs travaux constituent par là-même la clé du mystère que recouvre le dualisme entre le corps et l'esprit. Comme le montrent les travaux de Daniel Dennett (1991), une telle affirmation est bien sûr discutable⁽⁹⁾. Pour autant, aussi contestable soit-elle sur le plan scientifique, cette affirmation constitue un bon indicateur des niveaux de confiance dont font preuve les personnes auxquelles il faut se confronter lorsque l'on fait de l'intelligence artificielle. À l'instar d'Herbert A. Simon et d'Allen Newell, les pionniers de l'intelligence artificielle sont des chercheurs disposant d'importantes compétences dans le domaine des mathématiques, celui de l'informatique, mais aussi celui des sciences humaines et sociales. Partant, pour survivre dans cet environnement hyperconcurrentiel, chacun doit apprendre à faire valoir ses compétences et sa singularité de chercheur auprès des acteurs politiques et économiques. C'est pourquoi, afin d'impressionner les investisseurs tout en suscitant leur

⁽⁹⁾ Précisons toutefois qu'il ne faut pas, non plus, négliger le rôle fondamental qu'ont joué Herbert A. Simon et Allen Newell dans le développement de la philosophie de l'esprit.

pleine confiance⁽¹⁰⁾, les chercheurs en intelligence artificielle doivent afficher une assurance sans faille vis-à-vis de leurs travaux. Une bonne partie des grandes figures de l'intelligence artificielle est ainsi amenée, dès la fin des années 1950, à formuler collectivement des prédictions souvent optimistes. Encore une fois, Herbert A. Simon et Allen Newell font partie des chercheurs qui n'ont pas hésité à abuser de leur autorité scientifique pour donner de la crédibilité à des promesses démesurées :

- « 1. [... D']ici dix ans, un ordinateur numérique sera le champion du monde d'échecs, à moins que les règles ne l'excluent de la compétition.
- 2. [... D']ici dix ans, un ordinateur numérique découvrira et prouvera un nouveau théorème mathématique important.
- 3. [... D']ici dix ans, un ordinateur numérique écrira de la musique qui sera acceptée par les critiques comme ayant une valeur esthétique considérable.
- 4. [... D']ici dix ans, la plupart des théories en psychologie prendront la forme de programmes informatiques ou d'énoncés qualitatifs sur les caractéristiques des programmes informatiques » (traduit par nous ; Simon et Newell, 1958, pp. 7-8).

Bien sûr, les deux compères de Carnegie Mellon font preuve d'une certaine perspicacité : quelques-unes des prédictions susmentionnées se sont effectivement réalisées. Il n'en reste pas moins qu'au sens strict du terme, elles sont toutes fausses et auraient par conséquent mérité d'être modérées : par exemple, il faut attendre 1997 – et non pas 1968 – pour que « Deep Blue » gagne face à Gary Kasparov. Mais, pour Herbert A. Simon et Allen Newell, l'important n'est pas tellement que leurs anticipations soient vraies ou fausses. Les deux collègues ont rapidement compris que les applications de l'intelligence artificielle peuvent devenir un marché particulièrement juteux de technologies de gestion à destination de l'ensemble des acteurs engagés dans la chaîne de production et de distribution des biens et des services (consommateurs y compris ; Cochoy, Smolinski & Vayre, 2016). Pour Herbert A. Simon et Allen Newell, l'important est que leurs anticipations apparaissent comme autant de raisonnables et heureuses promesses pour le complexe militaro-industriel qu'ils connaissent bien. Car ces deux chercheurs savent comment fonctionnent les affaires, et plus particulièrement celles qui concernent les domaines militaire et industriel : rappelons qu'Herbert A. Simon et Allen Newell sont également consultants pour la RAND Corporation. En d'autres termes, s'ils ont conscience, en tant que chercheurs, que leurs prédictions n'ont pas valeur de vérité au sens scientifique du terme, Herbert A. Simon et Allen Newell savent, en tant que consultants, que celles-ci sont autant de promesses susceptibles de séduire les investisseurs.

⁽¹⁰⁾ Rappelons que les sommes en jeu sont immenses puisque, même dans les années 1960, elles peuvent déjà dépasser le million de dollars par an.

Des déceptions et des peurs

L'intelligence artificielle a toujours été une discipline qui dérange, dans la mesure où elle constitue une manière de saisir le comportement humain d'un point de vue à la fois objectif et distancié. Il faut en ce sens bien comprendre que lorsqu'Alan M. Turing soutient, en 1950, qu'une machine peut être capable de produire des pensées, c'est surtout pour bousculer l'*intelligentsia* du moment. À cette période, ce précurseur de l'intelligence artificielle sait déjà qu'il est homosexuel, et il accepte mal la *doxa* de son temps : il se méfie par conséquent des argumentations souvent religieuses, autoritaires et douteuses qui ont, par exemple, soutenu que les femmes, puis les animaux, ne pouvaient pas faire preuve d'intelligence (Turing, 1950). Pourquoi l'intelligence apparaît-elle comme si sacrée aux yeux de certains ? Qu'est-ce qui interdit de formuler l'hypothèse qu'une machine puisse réaliser des comportements intelligents ? Aussi, Herbert A. Simon et Allen Newell ont rapidement compris que cette provocation que compose, en lui-même, le programme de recherche en intelligence artificielle peut constituer un avantage socioéconomique. Ils savent que le marché aime l'innovation et que celui-ci peut par conséquent constituer un allié précieux pour lutter contre les traditions scientifiques qui pourraient enrayer le développement de l'intelligence artificielle. À tout le moins au départ et pour instituer cette discipline comme scientifique, les chercheurs qui l'exercent ne peuvent effectivement que partiellement s'appuyer sur les institutions académiques. Il leur faut trouver d'autres moyens de faire exister durablement l'intelligence artificielle. Mais Herbert A. Simon et Allen Newell ne sont pas les seuls à avoir saisi cela. Par exemple, Marvin L. Minsky participe, à sa façon, à diffuser auprès du grand public une représentation plus ou moins réaliste de ce que peut produire l'intelligence artificielle. Il le fait, notamment, par le biais de la science-fiction, par exemple, en conseillant Stanley Kubrick durant le tournage de *2001 l'Odyssée de l'espace* (1968 ; Ganascia, 2019 [2016]). Et il existe bien d'autres chercheurs ayant mobilisé d'autres façons encore de faire parler de l'intelligence artificielle. L'important, ici, est de garder en tête que l'objectif général de ces chercheurs est, de façon plus ou moins explicite, de favoriser l'expansion du marché de cette science en introduisant celle-ci dans le marché de la communication.

À partir de là, de nombreuses difficultés commencent à apparaître sur le plan socioéconomique. Les pionniers de l'intelligence artificielle développent probablement trop vite les voies de communication qu'ils instituent entre leur science et le marché. Cette précipitation s'explique, notamment, par le fait qu'ils ont un sentiment d'urgence, dans la mesure où ils ont besoin de disposer de machines puissantes et coûteuses pour dépasser leurs concurrents. Ainsi, si, par le biais du marché, la compétition scientifique est plus fortement influencée par la dynamique politique et économique du système capitaliste, il n'en reste pas moins que l'intelligence artificielle prend des forces et s'instaure solidement au sein des territoires académiques. Au début des années

1960, les promesses vont alors bon train. Les investisseurs se laissent séduire. Les médias s'emballent. C'est l'âge d'or de l'intelligence artificielle. Mais cette période phare ne dure pas. Après l'euphorie, les déceptions commencent et sont à la hauteur des espérances. Par exemple, à la suite des évaluations plutôt mauvaises exprimées par l'Automated Language Processing Advisory Committee (ALPCA ; Pierce *et al.*, 2019 [1966]) sur les avancées réalisées dans le champ de la traduction automatique, le gouvernement américain décide, dès 1966, de stopper les investissements qui devaient, au départ, leur permettre de traduire les communiqués de l'Union soviétique (Hutchins, 1996). Et, nous l'avons déjà mentionné, cet insuccès n'est que le premier d'une longue série : du point de vue des acteurs économiques et politiques, les échecs sont nombreux. Par exemple, le robot Shakey (Nilsson, 2019 [1984]) ne peut faire l'objet d'aucune exploitation militaire ou industrielle, étant donné que les tâches qu'il effectue sont lentes et prennent la forme d'un enchaînement saccadé de petites séquences. En outre, Shakey est très sensible aux changements de cadre : il suffit que ses roues patinent pour que la représentation qu'il se fait de son environnement ne soit plus adaptée à la situation réelle (Hart et Nilsson, 1972). Le « Speech Understanding Research » (SUR) de Donald E. Walker (2019 [1973]) constitue également un véritable fiasco sur le plan appliqué. Ce système n'est pas utilisable dans la mesure où ses utilisateurs doivent se soumettre à d'importantes contraintes grammaticales pour que SUR puisse traiter leur requête en temps réel. De telle sorte qu'au final, cette technologie est plus difficile à utiliser que les systèmes classiques de sélection dans un menu (Crevier, 1997 [1993]). Et il en va de même des systèmes experts qui, à tout le moins dans les années 1980, connaissent pourtant un véritable succès. C'est le cas de XCON (Bachant et McDermott, 1984). Au bout de quelques années, la mise à jour de sa base de connaissances est un véritable enfer. Pour reprendre les termes de Dominique Cardon et de ses collègues (2018), XCON devient une véritable « cathédrale de règles » : ce système expert est alors d'une telle complexité que la Digital Equipment Corporation (DEC) doit investir plus de deux millions de dollars par an pour réaliser sa maintenance (Simon, 1987).

Pour autant, l'intelligence artificielle ne fait pas que décevoir les investisseurs. Dès le départ, elle fait également peur aux citoyens-consommateurs. Pour bien comprendre cela, il faut d'abord saisir que le marché de la communication gravitant autour de l'intelligence artificielle est fructueux, notamment parce qu'il suscite des fantasmes, des promesses, mais aussi des risques plus ou moins avérés. Le problème est alors que, du point de vue du grand public, ce marché produit un mélange d'informations qui transforme l'intelligence artificielle en un concept fourre-tout et inquiétant. Comme le montrent les extraits d'interviews suivants, l'intelligence artificielle finit par faire peur aux citoyens-consommateurs. Cette peur peut être comprise de diverses façons.

Par exemple, pour Michel Melkanoff, elle n'est pas raisonnable dans le sens où le risque de l'intelligence artificielle ne vient pas des machines, mais des hommes qui les conçoivent et les utilisent :

- « Il y a des gens qui ont peur des machines [...] qui [...] deviendront des robots super-humains et qui vont conquérir le monde [...]. J'ai quelques mots à dire là-dessus. [...] [Personne ne] peut sérieusement avoir peur [...] d'une pile de fils de fer et de métal, [sinon] il n'est pas raisonnable. Intervieweur : "La bombe atomique est une pile de fils de fer et de métal !" Bien sûr, mais ce n'est pas de la bombe qu'on a peur. Ce sont des gens qui peuvent la lancer. Dans ce sens aussi, il y a peut-être danger quant aux hommes qui peuvent se servir de l'ordinateur » (Melkanoff, cité in Moreuil, 1972).

Abraham Moles pense, quant à lui, un peu différemment. Pour ce spécialiste des sciences de l'information et de la communication, cette peur découle plutôt de ce qu'il appelle « une inquiétude sociologique », c'est-à-dire d'une crainte plus ou moins rationnelle de la population concernant les formes d'aliénation que les applications de l'intelligence artificielle peuvent permettre :

- « Le public a peur des machines [...] au fur et à mesure qu'elles se révèlent [...] et [...] s'insinuent dans notre vie quotidienne [...]. René de Possel faisait remarquer que le jour où les quarante-sept millions de Français seront mis avec mille ou deux mille critères chacun dans une carte perforée, plus de police, plus de procès-verbaux, etc. Tout au fichier central, et on ressort l'individu qui n'est plus anonyme. Il est [...] personnalisé. Il ne peut pas se reposer sur la liberté interstitielle, sur le jeu des institutions. Il est prisonnier ! Et je crois que ceci est la raison [...] de la peur » (Moles, cité in Lallier, 1963).

En référence aux travaux de Madeleine Akrich, de Michel Callon et de Bruno Latour (2006), c'est donc par le biais d'une pratique économique plus ou moins maîtrisée que les chercheurs en intelligence artificielle traduisent ce qu'ils font en promesses et en applications destinées à intéresser et à enrôler les investisseurs. Ces traductions sont ensuite amplifiées par le marché de la communication qui gravite autour de l'intelligence artificielle. Le problème est alors que, dès le début, cette amplification est aussi une déformation obscurcissant ce qu'est l'intelligence artificielle. Le grand public finit par oublier durablement que l'intelligence artificielle est d'abord une science. Et il n'est pas impossible qu'il en soit de même encore aujourd'hui : qui a une idée des programmes de recherche en intelligence artificielle, ou encore des enjeux épistémologiques, sociaux et humains qui y sont associés ? Ainsi, des années 1960 aux années 1980, pour le grand public, l'intelligence artificielle est, au mieux, une sorte de mécanisme ou d'énergie, à la fois flou(e) et insaisissable, qui existe à l'intérieur des machines pour réguler leur fonctionnement et qui grandit plus ou moins dangereusement... Au pire, c'est une immense arnaque !

Des critiques pour rassurer et bousculer le marché

Pour une grande partie de la population, le concept d'intelligence artificielle perd de son sens. Il devient alors embarrassant. Mais cet embarras ne gêne pas tellement les chercheurs : même ceux qui n'aiment pas trop cette notion s'y accommodent assez facilement (voir la section « Des styles de recherche différents »). Ce sont surtout les industriels engagés dans le développement de l'informatique qui sont embarrassés par ce concept. C'est par exemple le cas d'IBM :

- « Ces projets d'IA menés au sein de la firme [IBM] furent finalement victimes de leur propre succès. [...] Thomas J. Watson fut sommé durant une réunion d'actionnaires d'expliquer pourquoi la compagnie consacrait les dollars de la recherche à des sujets aussi frivoles. De surcroît, le département marketing d'IBM notait alors un changement alarmant dans la psychologie du consommateur : celui-ci voyait l'ordinateur comme une menace et, par crainte, s'en détournait. [...] Pour Watson, la coupe était pleine [...]. Les futures campagnes de marketing de la firme [...] balayeraient l'image héritée de la science-fiction d'un ordinateur agissant comme un gigantesque cerveau, pour la remplacer par le cliché rassurant d'une machine maniant simplement des nombres. Les ordinateurs, clameraient inlassablement IBM, [...] ne feraient que ce qu'on leur dirait. Jamais ils ne voleraient la place d'un dirigeant, leur seul talent résidant dans leur rapidité à traiter des flots et des flots de chiffres » (Crevier, 1997 [1993], p. 78).

L'embarras qu'éprouvent les industriels vis-à-vis de l'intelligence artificielle est alors amplifié par les multiples déceptions dont nous avons parlé. De telle sorte que, nous l'avons également signalé, les acteurs économiques et politiques sont conduits à remettre en question leur engagement dans le développement de cette discipline. Dès lors, ils commencent à porter un réel intérêt aux critiques de l'intelligence artificielle. Stuart E. Dreyfus, un consultant de la RAND Corporation, en profite alors pour mettre son frère, Hubert L. Dreyfus, philosophe, en contact avec cette organisation. Ce dernier est de cette manière sommé de déterminer, sur le plan philosophique, la validité du projet de l'intelligence artificielle : la RAND Corporation souhaite qu'il prédise la capacité de cette science à valider l'hypothèse selon laquelle il est possible de reproduire matériellement des comportements que les hommes considèrent comme intelligents. Et, pour Hubert L. Dreyfus (1972), la réponse est évidemment négative. Selon lui, l'intelligence n'a rien à voir avec un système de computation de représentations symboliques et ne consiste pas à effectuer des opérations logiques. Hubert L. Dreyfus pense plus exactement qu'à la différence des humains, les machines calculent de manière brutale : elles ne savent pas discriminer ce qui est ou non de l'ordre du pertinent (c'est le problème de la restriction). Pour le philosophe, si les humains savent traduire la complexité du monde en des réactions simples, il faut réduire, formaliser et expliciter cette même complexité pour qu'une machine

fasse de même. Or, pour Hubert L. Dreyfus, ceci est impossible à réaliser, à tout le moins pour l'intelligence artificielle symbolique qui domine à ce moment-là. Il ajoute, en ce sens, que, si les humains n'ont aucun mal à s'adapter à un changement d'environnement, ce n'est pas le cas des machines qui ne savent pas faire autrement que suivre des règles explicites (c'est le problème du cadre). En somme, pour Hubert L. Dreyfus, l'intelligence artificielle ne peut pas être considérée comme une science compte tenu de l'irrationalité de l'hypothèse de recherche qui est à son fondement. L'intelligence artificielle est, pour lui, une alchimie qui promet de décrocher la lune ! De manière plus générale, *Alchemy and artificial intelligence* (Dreyfus, 2019 [1965]) déborde alors la critique en règle : c'est un rapport provocant et mal documenté, qui appelle à la bagarre de rue. De telle sorte qu'avec ce document, la critique perd de son élégance et de sa pertinence. Même Joseph Weizenbaum, un des rares chercheurs en intelligence artificielle à s'être rangé dans le clan d'Hubert L. Dreyfus, pense que le rapport de son collègue est mauvais, notamment parce qu'il dénote un terrible manque de connaissance concernant le fonctionnement des ordinateurs (Crevier, 1997 [1993]).

Il n'en reste pas moins qu'en sollicitant les services de Hubert L. Dreyfus, la RAND Corporation lui donne du crédit et participe de cette manière à transformer le débat scientifique en une controverse socio-économique, qui prend rapidement les allures d'une dispute de comptoir. Au fur et à mesure des discussions, les arguments scientifiques sont remplacés par des insultes peu élégantes. Par exemple, à la provocation d'Hubert L. Dreyfus selon laquelle un enfant de six ans peut battre n'importe quel programme d'échecs – ce qui, à un moment donné, est vrai –, Seymour A. Papert répond en mettant publiquement le philosophe au défi de faire mieux que cet enfant⁽¹¹⁾ ! Bien sûr, Hubert L. Dreyfus a raison sur bien des aspects, et ses critiques sont, dans le fond, intéressantes : s'il a une mauvaise compréhension de l'intelligence artificielle, il a une bonne connaissance de la philosophie. Et c'est d'ailleurs cette connaissance qui lui permet de devancer John McCarthy en identifiant, bien avant lui, les deux grandes limites de cette discipline, c'est-à-dire le problème du cadre et celui de la restriction dont nous venons de parler. Mais peu importe qui a tort ou raison. Ici, l'important est de comprendre qu'à partir du rapport *Alchemy and artificial intelligence* (Dreyfus, 2019 [1965]), la controverse autour de l'intelligence artificielle déborde du champ scientifique pour se transformer en un problème socioéconomique.

D'une manière générale, ce problème recouvre deux grandes dimensions. La première correspond à une forme collective d'escalade d'engagements (Joule et Beauvois, 2002). Elle renvoie à l'idée que certains chercheurs, probablement trop engagés dans le développement de l'intelligence artificielle,

⁽¹¹⁾ Pour plus de détails sur les formes concrètes de cette dispute, nous renvoyons au rapport que Seymour A. Papert (2019 [1968]) a rédigé en réponse à *Alchemy and artificial intelligence* (Dreyfus, 2019 [1965]).

continuent à s'accrocher à leurs promesses (notamment afin de ne pas perdre la face) tout en cherchant à regagner la confiance du marché (et plus particulièrement celle des investisseurs). La deuxième dimension constitue une forme de reconstruction fictive de la réalité. Elle rassemble les acteurs de la critique qui souhaitent jouer le jeu d'industriels comme IBM en niant l'existence de l'intelligence artificielle. L'ennui est alors que ces acteurs oublient, plus ou moins volontairement, une chose primordiale : l'intelligence artificielle n'est pas une forme de réflexion machinique et autonome que pourraient développer les technologies issues de l'informatique. C'est une science dont il est difficile de nier l'existence. Le problème socio-économique de l'intelligence artificielle est donc double. D'une part (pour l'escalade d'engagements), il est associé au fait que le financement de cette science repose sur des coups de bluff qui risquent de perpétuer les cycles de promesses/déceptions pour finalement perturber les bonnes réalisations des programmes scientifiques de cette discipline. D'autre part (pour la reconstruction fictive de la réalité), il est lié au fait que la stratégie qui consiste à déformer ce qu'est l'intelligence artificielle pour mieux nier son existence constitue une façon de rendre invisible son développement. Cette invisibilisation confère alors un important pouvoir aux marchands : sans même en avoir réellement conscience, ils deviennent les seuls à assurer et à contrôler la diffusion des applications technologiques issues de l'intelligence artificielle, mais aussi, par conséquent, les formes de son financement.

Après la prédominance, pendant plusieurs années, du problème de l'escalade d'engagements (cf. les hivers de l'intelligence artificielle), c'est la deuxième dimension du problème socio-économique de l'intelligence artificielle (cf. la reconstruction fictive de la réalité dont nous venons de parler) qui, à partir des années 1980, prend le dessus. Comme l'illustre la Figure 2, durant cette période, les entreprises du secteur de l'informatique ne veulent plus, en effet, parler de systèmes experts, et encore moins d'intelligence artificielle. Elles préfèrent prendre les habits de simples fournisseurs de solutions informatiques, afin d'apparaître, aux yeux de leurs clients, comme des entreprises à la fois clairvoyantes et consciencieuses. Les programmes d'intelligence artificielle deviennent de cette façon des sortes d'applications cachées : ils sont discrètement intégrés aux programmes informatiques plus classiques. Patrick H. Winston, par exemple, connaît bien cette stratégie. Dans les années 1980, comme la plupart de ses collègues, il possède une entreprise de conception de programmes informatiques, qu'il diffuse déjà selon un système qu'il qualifie de « pain aux raisins » :

- « L'IA est à présent intégrée dans les systèmes comme les raisins dans le pain : elle n'occupe pas nécessairement beaucoup de volume, mais peut avoir à elle seule beaucoup de qualités nutritives. Vous ne pouvez pas dissocier les raisins du pain ; et il existe différentes variétés de raisins » (Winston, cité in Crevier, 1997 [1993], p. 252).

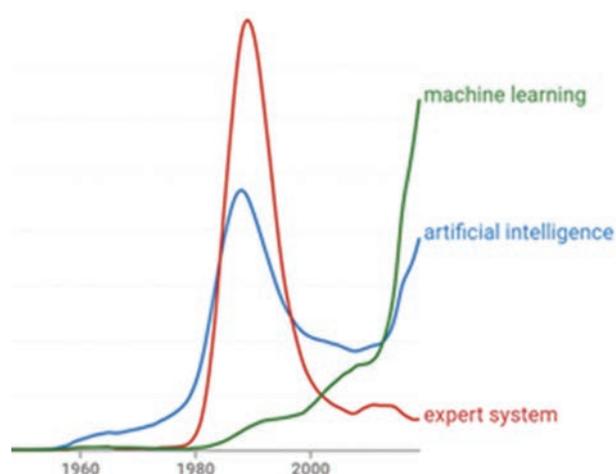


Figure 2. Représentation graphique du nombre d'occurrences des appellations "artificial intelligence", "expert system" et "machine learning" sur Google Books Ngram Viewer⁽¹²⁾

Le cas typique du système « pain aux raisins » est celui du programme d'assistance commerciale. Classiquement, un tel programme se contente de contrôler la disponibilité du produit, d'enregistrer la transaction, de faire la facture et d'informer le service en charge de l'expédition. Ce programme peut alors être augmenté d'un système expert destiné, par exemple, à suggérer à l'utilisateur des produits de remplacement en cas de rupture de stock. Partant, l'avantage de cette nouvelle manière de penser le modèle d'affaires de l'intelligence artificielle est qu'en rendant invisible l'existence de celle-ci le battage médiatique que provoque cette science s'éteint progressivement. La confusion sur ce qu'est l'intelligence artificielle disparaît en même temps que les inquiétudes, les peurs et les risques associés. Il y a toutefois un inconvénient : cette discipline et les applications technologiques qu'elle fabrique continuent en réalité d'exister. Comme nous l'avons déjà dit, le souci tient alors au fait que l'organisation de la diffusion des applications technologiques de l'intelligence artificielle bascule entre les seules mains d'acteurs économiques dont les intérêts ne peuvent se confondre avec ceux de la société. Et il faudra attendre l'augmentation importante et rapide que nous avons connue ces dix dernières années en termes de capacités de production, de stockage, de traitement des données numériques, et qui ouvre la voie à un nouvel âge de l'apprentissage artificiel⁽¹³⁾ (*machine learning* ; voir Figure 2), pour que la société prenne conscience de ce problème et s'interroge à nouveau sur les enjeux économiques, sociaux

⁽¹²⁾ Notons que les courbes présentées dans cette figure constituent une sorte de caisse de résonance de la réalité empirique et sont par conséquent en léger décalage avec celles que nous présentons dans cet article, compte tenu qu'elles sont produites à partir de l'analyse statistique du corpus de textes du service Google Books. Par exemple, le pic d'occurrences de l'appellation « expert system » se réalise en 1988 alors que l'âge d'or des systèmes experts aux États-Unis se situe plutôt entre la fin des années 1970 et le début des années 1980 (Crevier, 1997 [1993]).

⁽¹³⁾ Avec des réussites fortes comme les programmes « AlphaGo » de DeepMind, « Watson » de IBM, ou encore « DeepL Translator » de DeepL.

et humains associés au développement de l'intelligence artificielle et à la diffusion des applications qu'elle fabrique (cf. le mouvement *big data* ; Cardon, 2015 ; Vayre, 2016).

Conclusion

L'intelligence artificielle est une discipline scientifique dont le programme de recherche est, à tout le moins à l'origine, fortement expérimental : il s'agit de tester l'hypothèse selon laquelle il est possible de reproduire matériellement l'intelligence humaine et, plus largement, celle de l'ensemble des êtres vivants. L'histoire a alors montré que, sur le plan purement scientifique, l'intelligence artificielle a eu le mérite de mettre en question ce qu'est l'intelligence et de susciter ainsi des avancées importantes dans le domaine des sciences cognitives, mais aussi dans celui des sciences humaines et sociales. Nous avons ajouté que faire de l'intelligence artificielle ne se limite pas à effectuer une activité scientifique, c'est aussi une pratique économique. Pour exercer cette science, il faut attirer l'attention des investisseurs afin de pouvoir disposer d'équipements qui coûtent chers : dans le domaine de l'intelligence artificielle, science et marché sont intimement intriqués. Aussi, dans un contexte socio-économique où les grands acteurs du numérique (les GAFAM pour Google, Apple, Facebook, Amazon et Microsoft) tendent à attirer vers eux les plus importants chercheurs en intelligence artificielle, nous pensons qu'il n'est pas inutile de rappeler qu'historiquement l'intelligence artificielle est d'abord une discipline scientifique expérimentale, dont la finalité est de mieux comprendre ce qu'est l'intelligence et comment il est ou non possible de la matérialiser. Nos travaux ont en ce sens l'avantage de rappeler que, pour les pionniers, l'intelligence artificielle n'est pas un produit de consommation donnant lieu de développer des services automatisés comme, c'est par exemple le cas à l'heure actuelle, avec l'économie des plateformes : c'est une science qui permet de poser des questions de fond pouvant potentiellement déboucher sur des applications heureuses⁽¹⁴⁾. En revanche, nos études nous rappellent que, dès le début, l'intelligence artificielle est également une pratique économique qui consiste à habiller politiquement et économiquement des idées scientifiques de façon à les attacher à des usages sociotechniques. Nous avons alors cherché à montrer comment cet habillage impulse historiquement un jeu d'attraction et de répulsion vis-à-vis des acteurs économiques (des investisseurs et des consommateurs), qu'il convient de mieux comprendre, notamment si l'on souhaite mieux saisir et maîtriser la réalité des enjeux associés au développement de l'intelligence artificielle.

⁽¹⁴⁾ Même si, dans les faits et au départ, ces applications ont été plutôt malheureuses.

Références

- AKRICH M. (1989), « La construction d'un système socio-technique. Esquisse pour une anthropologie des techniques », *Anthropologie et Sociétés*, 13(2), pp. 31-54.
- AKRICH M., CALLON M. & LATOUR B. (2006), *Sociologie de la traduction : textes fondateurs*, Presses des Mines, Paris.
- BACHANT J. & McDERMOTT J. (1984), "R1 revisited: Four years in the Trenches", *AI Magazine*, 5(3), pp. 21-32.
- BOURDIEU P. (1976), « Le champ scientifique », *Actes de la recherche en sciences sociales*, 2(2-3), pp. 88-104.
- BRUSCAGLIONI L. (2016), "Theorizing in grounded theory and creative abduction", *International Journal of Methodology*, 50(5), pp. 2009–2024.
- BUCHANAN B. G. & FEIGENBAUM E. A. (1978), "Dendral and meta-dendral: their applications dimension", *Artificial Intelligence*, 11(1-2), pp. 5-24.
- BUCHANAN B. G. & SHORTLIFFE H. E. (1984), *Rule-based expert systems: the mycin experiments of the stanford heuristic programming project*, Addison Wesley, Boston.
- BURGUIÈRE A. (1979), « Les Annales, 1929-1979 », *Annales. Économies, sociétés, civilisations*, n°6, pp. 1344-1346.
- CARDON D. (2015), *À quoi rêvent les algorithmes : nos vies à l'heure des big data*, Paris, Seuil.
- CARDON D., COINTET J.-P. & MAZIÈRES A. (2018), « La revanche des neurones : l'invention des machines inductives et la controverse de l'intelligence artificielle », *Réseaux*, 5(211), pp. 173-220.
- CHANGEUX J.-P. (1983), *L'Homme neuronal*, Paris, Fayard.
- CHOMSKY N. (1965), *Aspects of the theory of syntax*, MIT Press, Cambridge.
- COCHOY F., SMOLINSKI J. & VAYRE J.-S. (2016), "From marketing to 'market-things' and 'market-Iting': Accounting for technicized and digitalized consumption", in CZARNIAWSKA B. (dir.), *A research agenda for management and organization studies*, Edward Elgar, Glasgow, pp. 26-37.
- CREVIER D. (1997 [1993]), *À la recherche de l'intelligence artificielle*, Paris, Flammarion.
- Davis R. (1978), "Knowledge acquisition in rule-based systems – knowledge about representations as a basis for system construction and maintenance", *Pattern-Directed Inference Systems*, pp. 99-134.
- DENNETT D. (1991), *Consciousness explained*, Little, Brown and Co, Boston.
- DREYFUS H. L. (1972), *What computers can't do: a critique of artificial reason*, Harper & Row, New York.
- DREYFUS H. L. (2019 [1965]), *Alchemy and artificial intelligence*, en ligne : <https://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf>
- FALETTI J. (1982), "PANDORA - A program for doing common-sense planning in complex situations", *AAAI-82 Proceedings*, pp. 185-188.
- FLAMM K. (1987), *Targeting the computer: government support and international competition*, Brookings Institution, Washington.
- FODOR J. & PYLYSHYN Z. W. (1988), "Connectionism and cognitive architecture: A critical analysis", *Cognition*, 28(1-2), pp. 3-71.
- GANASCIA J.-G. (2019 [2016]), « Marvin Minsky : un des cerveaux de l'intelligence artificielle », en ligne : <https://interstices.info/marvin-minsky-un-des-cerveaux-de-lintelligence-artificielle/>
- GELERTNER H. L. & ROCHESTER N. (1958), "Intelligent behavior in problem-solving machines", *IBM Journal of Research and Development*, 2(4), pp. 336-345.
- GINZBURG C. (1980), « Signes, Traces, Pistes – Racines d'un paradigme de l'indice », *Le Débat*, 6(6), pp. 3-44.
- HART P. & NILSSON N. (1972), *Shakey: experiments in robot planning and learning*, en ligne : <https://www.youtube.com/watch?v=GmU7SimFkpU>

- HEIMS S. J. (1991), *The cybernetics group*, MIT Press, Cambridge.
- HODDESON L., HENRIKSEN P. W., MEADE R. A. & WESTFALL C. (1992), *Critical assembly: A technical history of Los Alamos during the Oppenheimer years, 1943–1945*, Cambridge University Press, Cambridge.
- HODGES A. (2015 [1983]), *Alan Turing : le génie qui a décrypté les codes secrets nazis et inventé l'ordinateur*, Paris, Payot.
- HUTCHINS J. (1996), "ALPAC: the (in)famous report", *MT News International*, n°14, pp. 9-12.
- JOULE R.-V. & BEAUVOIS J.-L. (2002), *Petit traité de manipulation à l'usage des honnêtes gens*, Grenoble, Presses Universitaires de Grenoble.
- JULIA L. (2019), *L'intelligence artificielle n'existe pas*, Paris, First.
- LALLIER J. (1963), *Visa pour l'avenir : des machines à penser ?*, en ligne : <https://www.ina.fr/video/CPF86656441/des-machines-a-penser-video.html>
- LATOURET B. (1987), *Science in action: How to follow scientists and engineers through society*, Harvard University Press, Cambridge.
- LENAT D. B. (1977), "Automated theory formation in mathematics", *Proceedings of the 5th international joint conference on Artificial intelligence*, vol. 2, pp. 833-842.
- LENAT D. B. (1983), "Eurisko: a program that learns new heuristics and domain concepts", *Artificial Intelligence*, 21(1-2), pp. 61-98.
- MCCARTHY J., MINKSY M. L., ROCHESTER N., SHANNON C. E. (2019 [1955]), *A proposal for the Dartmouth summer research projet on artificial intelligence*, en ligne : <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>
- McDERMOTT J. (1982), "R1: A rule-based configurer of computer systems", *Artificial Intelligence*, 19(1), pp. 39-88.
- MINSKY M. L. (1954), *Theory of neural-analog reinforcement systems and its application to the brain model problem*, Princeton University, Princeton.
- MINSKY M. L. (1974), *A framework for representing knowledge*, Massachusetts Institute of Technology, Cambridge.
- MINSKY M. & PAPER S. A. (1969), *Perceptrons*, MIT Press, Cambridge.
- MOREUIL F. (1972), *Un certain regard - Les machines et les hommes : l'intelligence artificielle*, en ligne : <https://www.ina.fr/video/CPF86655105/l-intelligence-artificielle-video.html>
- MOSES J. (1967), *Symbolic integration*, Massachusetts Institute of Technology, Cambridge.
- MOUNIER-KUHN P.-E. (2010), *L'informatique en France de la seconde Guerre Mondiale au Plan Calcul. L'émergence d'une science*, Paris, Presses de l'Université Paris-Sorbonne.
- NEWELL A. & SIMON H. A. (1956), "The logic theory machine: a complex information processing system", *IRE Transactions on Information Theory*, 2(3), pp. 61-79.
- NEWELL A., SHAW J. C. & SIMON H. A. (1959), "Report on a general problem-solving program", *Proceedings of the International Conference on Information Processing*, pp. 256-264.
- NILSSON N. J. (2019 [1984]), *Shakey the robot*, en ligne : <http://www.ai.sri.com/pubs/files/629.pdf>
- NORBERG A. L. (2019 [1989]), *An interview with Marvin L. Minsky*, en ligne : <https://conservancy.umn.edu/bitstream/handle/11299/107503/oh179mim.pdf?sequence=1&isAllowed=y>.
- ORLÉAN A. (2000), « L'individu, le marché et l'opinion : réflexions sur le capitalisme financier », *Esprit*, 11(269), pp. 51-75.
- PAPER S. A. (2019 [1968]), *The artificial intelligence of Hubert L. Dreyfus: a budget of fallacies*, en ligne : <https://dspace.mit.edu/bitstream/handle/1721.1/6084/AIM-154.pdf?sequence=2&isAllowed=y>
- PAPER S. A. (1971), *A computer laboratory for elementary schools*, Massachusetts Institute of Technology, Cambridge.
- PASSERON J.-C. (1991), *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*, Paris, Nathan.
- PIERCE J. R., CARROLL J. B., HAMP E. P., HAYS D. G., HOCKETT C. F., OETTINGER A. G. & PERLIS A. (2019 [1966]), *Language and machine: computers on translation and linguistics*, en ligne : <http://www.mt-archive.info/ALPAC-1966.pdf>
- PRATT V. (1995 [1987]), *Machine à penser : une histoire de l'intelligence artificielle*, Paris, Presses Universitaires de France.
- ROSE F. (1986 [1984]), *L'intelligence artificielle : histoire d'une recherche scientifique*, Paris, Payot.
- ROSENBLATT F. (1958), "The Perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, 65(6), pp. 386-408.
- SCHANK R. C. (1978), "Computer understanding of natural language", *Behavior Research Methods & Instrumentation*, 10(2), pp. 132–138.
- SCHANK R. C. & ABELSON R. P. (1977), *Scripts, plans, goals, and understanding*, Lawrence Erlbaum, Hillsdale.
- SEARLE J. R. (2002), *Consciousness and language*, Cambridge University Press, Cambridge.
- SELFRIDGE O. G. (1959), "Pandemonium: A paradigm for learning", *Proceedings of the Symposium on Mechanisation of Thought Processes*, pp. 511-529.
- SIMON H. A. (1945), *Administrative behavior: A study of decision making processes in administrative organization*, The Free Press, New York.
- SIMON H. A. (1991), *Models of my life*, Basic Books, New York.
- SIMON H. A. & NEWELL A. (1958), "Heuristic problem solving: The next advance in operations research", *Operations Research*, 6(1), pp. 1-10.
- SIMON R. (1987), "The morning after", *Forbes*, 140(8), pp. 164-168.
- SLAGLE J. R. (1961), *A heuristic program that solves symbolic integration problems in freshman calculus: Symbolic automatic integrator (SAINT)*, Massachusetts Institute of Technology, Cambridge.
- TAYLOR W. K. (1956), "Electrical simulation of some nervous system functional activities", *Information theory*, n°3, pp. 314-328.
- THOUARD D. (2007), *L'interprétation des indices. Enquête sur le paradigme indiciaire avec Carlo Ginzburg*, Villeneuve-d'Ascq, Presses Universitaire du Septentrion.
- TURING A. M. (1950), "Computing machinery and intelligence", *Mind*, vol. 59, pp. 433-460.
- VAYRE J.-S. (2016), *Des machines à produire des futurs économiques : sociologie des intelligences artificielles marchandes à l'ère du big data*, Toulouse, Université Toulouse Jean Jaurès.
- WALKER D. E. (2019 [1973]), *Speech understanding research*, en ligne : <https://apps.dtic.mil/dtic/tr/fulltext/u2/757049.pdf>
- WIDROW B. & HOFF M. E. (1960), "Adaptive switching circuits", *1960 IRE WESCON Convention Record*, n°4, pp. 96-104.
- WILENSKY R. (1977), "PAM - A Program That Infers Intentions", *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, vol. 1, pp. 1-15.
- WINOGRAD T. (1972), "Understanding natural language", *Cognitive Psychology*, 3(1), pp. 1-191.
- WINSTON P. H. (1970), *Learning structural descriptions from examples*, Massachusetts Institute of Technology, Cambridge.