

Université de Montréal

Analyse des représentations sociales du  
concept « d'intelligence » dans les discours sur  
l'Intelligence Artificielle.

Par  
Adèle Ghiringhelli

Département de Sociologie

Faculté des Arts et Sciences

Mémoire présenté à la Faculté des Études Supérieures et Postdoctorales  
en vue de l'obtention du grade de maîtrise  
en sociologie.

Août 2019

© Adèle Ghiringhelli, 2019

Université de Montréal  
Faculté des Études Supérieures et Postdoctorales

---

*Ce mémoire intitulé*

**Analyse des représentations sociales du concept « d'intelligence » dans les discours sur l'Intelligence Artificielle.**

*Présenté par*

**Adèle Ghiringhelli**

*A été évalué par un jury composé des personnes suivantes*

**Jacques Hamel**  
Président du jury

**Céline Lafontaine**  
Directeur de recherche

**Nicolas Le Dévédec**  
Membre du jury

## Résumé

Technologie en vogue pour les uns et domaine de recherche scientifique pour les autres, l'Intelligence Artificielle (IA) constitue avant tout un projet idéologique : celui de façonner une créature artificielle « à l'image de l'Homme » (Breton, 1995). Toutefois, l'engouement médiatique face à l'IA retranscrit un débat exclusivement centré sur les possibilités qu'elle génère et les enjeux qu'elle soulève. Essentielle à la compréhension du phénomène social, l'étude sociologique de l'évolution de ce projet idéologique constitue un manque auquel ce mémoire entend combler. Il s'agit d'établir sur quelle conception de l'être humain et de son intelligence s'est développé ce projet. Ce travail de recherche porte sur les représentations de l'intelligence véhiculées par les discours médiatiques centrés sur l'IA. Analyser les représentations sociales de l'intelligence dans une perspective constructiviste permet non seulement de mieux appréhender le phénomène social qu'est l'IA, mais nous éclaire par-dessus tout sur la vision de l'être humain qu'elle suppose. Nous tenterons de démontrer en quoi l'IA s'est développée sur la base d'une représentation réductrice de l'Homme et de son intelligence, sans jamais que les fondements de cette assise ne soient évoqués tant ils sont dissimulés aux confins des représentations sociales. D'autant que ce « paupérisme épistémique » (Robillard, 2019 : 3) n'est pas sans effet : de réels choix scientifiques en découlent entraînant la mise sur le marché d'applications concrètes qui guident et transforment notre quotidien. C'est pourquoi une réflexion de fond est de rigueur. Ce mémoire rend ainsi compte des intérêts sociaux affectés par la rationalité instrumentale technocratique.

**Mots-clés** : intelligence, Intelligence Artificielle, représentations sociales, technosciences, constructivisme, transhumanisme, vulgarisation, réductionnisme.

## **Abstract**

Popular technology for some and field of scientific research for others, Artificial Intelligence (AI) is an ideological project above all: shaping an artificial creature "in the image of Man" (Breton, 1995). However, the media's enthusiasm for AI reflects a debate that focuses exclusively on the opportunities it generates and the issues it raises. Essential to the understanding of the social phenomenon, the sociological study of the evolution of this ideological project constitutes a gap that this thesis intends to fill. The aim is to establish on what conception of the human being and his intelligence this project has developed. This research work focuses on the representations of intelligence conveyed by media discourses focused on AI. Analyzing the social representations of intelligence from a constructivist perspective not only allows us to better understand the social phenomenon that is AI, but also sheds light on the vision of the human being that it implies. We will try to demonstrate how AI has been developed on the basis of a reductive representation of Man and his intelligence without ever mentioning the foundations hidden on the edges of social representations. Especially since this "epistemic pauperism" (Robillard, 2019: 3) is not without effect: real scientific choices result from it, leading to the launch of concrete applications that guide and transform our daily lives. That is why a thorough reflection is required. This thesis thus aims to point out social interests affected by technocratic instrumental rationality.

**Keywords** : intelligence, Artificial Intelligence, social representations, technoscience, constructivism, transhumanism, popularization, reductionism.

## Table des matières

<b>Résumé .....</b>	<b>i</b>
<b>Remerciements .....</b>	<b>vi</b>
<b>Introduction.....</b>	<b>1</b>

### Chapitre I.

<b>Retour sociohistorique de l'Intelligence Artificielle .....</b>	<b>6</b>
<b>1. L'IA : une créature artificielle .....</b>	<b>7</b>
1.1. ... à l'image de l'Homme .....	7
1.2. ... au sein du modèle informationnel.....	11
<b>2. L'IA : du projet idéologique à la discipline scientifique .....</b>	<b>15</b>
2.1. Légitimation d'un fantasme.....	15
2.1.1. Les prémices .....	15
2.1.2. L'IA symbolique .....	19
2.2. Concrétisation technologique.....	24
2.2.1. Renaissance de l'IA .....	24
2.2.2. L'apprentissage profond : retour du connexionnisme .....	28

### Chapitre II.

<b>La notion d'intelligence.....</b>	<b>31</b>
<b>1. L'intelligence : un concept construit... ..</b>	<b>32</b>
1.1. ... à la lumière des représentations sociales .....	33
1.2. ...au fil des grandes approches historiques.....	35
1.2.1. L'intelligence hylémorphique d'Aristote.....	35
1.2.2. L'intelligence sacrée de Descartes .....	35
1.2.3. L'intelligence biologique de Darwin .....	36
1.2.4. L'intelligence informationnelle de Wiener .....	40
<b>2. La représentation de l'intelligence en IA.....</b>	<b>41</b>
2.1. Une définition en évolution .....	41
2.1.1. La rétroaction .....	41
2.1.2. La résolution de problème .....	43
2.1.3. L'optimisation.....	46
2.2. Une représentation réductrice de l'Homme .....	50
<b>3. Artificialisation de l'intelligence humaine et naturalisation de l'IA... ..</b>	<b>56</b>
3.1. ... à travers le reversement du dualisme nature-artifice. ....	56
3.2. ... à travers l'idéologie transhumaniste .....	60

## Chapitre III.

<b>La démarche méthodologique.....</b>	<b>65</b>
<b>1. L'objet de recherche .....</b>	<b>65</b>
1.1. L'Intelligence Artificielle en tant que phénomène social.....	65
1.2. La définition de l'objet de recherche : l'intelligence en IA.....	67
<b>2. L'analyse de discours.....</b>	<b>68</b>
2.1. Méthode de l'analyse discursive .....	69
2.2. La vulgarisation scientifique.....	70
2.3. L'aide à l'analyse qualitative via le logiciel Atlas.ti.....	72
<b>3. Construction des données sociologiques .....</b>	<b>73</b>
3.1. Identification du matériau d'analyse.....	73
3.2. Description des documents analysés.....	78
3.3. Mise en relation de l'objet de recherche et des matériaux.....	81

## Chapitre IV.

<b>La représentation de l'intelligence à travers les discours centrés sur l'IA..</b>	<b>83</b>
<b>1. Définition de l'intelligence dans les discours contemporains.....</b>	<b>83</b>
1.1. De la rétroaction à l'optimisation .....	84
1.2. L'optimisation.....	88
1.2.1. Données : nouveau modèle de représentation du monde.....	95
1.2.2. Recherche d'applications et fin de la théorie .....	99
1.2.3. Remise en question de la définition informationnelle de l'intelligence. ...	103
<b>2. Le réductionnisme.....</b>	<b>105</b>
2.1. Indifférenciation ontologique.....	108
2.2. Les qualités non modélisables intrinsèques à l'humain .....	112
<b>3. Entre nature et artifice .....</b>	<b>113</b>
3.1. Créature artificielle à l'image de l'Homme .....	113
3.2. Renversement du dualisme .....	115
3.2.1. Artificialisation de la nature.....	117
3.2.2. Naturalisation de la technique.....	124
<b>4. Considérations finales en vue des résultats de l'analyse discursive. ....</b>	<b>130</b>
<b>Conclusion .....</b>	<b>133</b>
<b>Bibliographie .....</b>	<b>135</b>

*À mes sœurs,  
dont l'intelligence dépasse l'entendement.*

# Remerciements

Je tiens tout d'abord à remercier ma directrice de recherche, Céline Lafontaine. Ses commentaires et ses encouragements auront été d'une profonde aide. Tout en me guidant à travers ce projet, elle a su stimuler ma réflexion en m'encourageant à aller toujours plus loin, sans quoi cette expérience n'aurait été aussi enrichissante. Mon esprit critique s'est forgé à travers les valeurs humanistes qui la guide et qu'elle transmet avec une grande passion ; pour cela je lui en suis infiniment reconnaissante.

Je voudrais aussi remercier ma mère pour le soutien à travers les nombreuses et répétitives relectures, et envers tous mes choix quoiqu'ils fussent. Les mots ne suffisent pas à exprimer toute la gratitude que j'ai envers l'amour inconditionnel que tu me portes et qui m'a permis d'avancer tout au long de cette expérience le cœur léger.

Un grand merci à mon père à qui je dois ma curiosité intellectuelle et dont l'approbation constitue, je le crains, ma motivation première.

Mes sœurs Clara, Marie, Anna ; merci d'être les plus magnifiques références dont une personne puisse rêver. Les savoir à mes côtés tout en résidant à l'autre bord de l'Atlantique est un immense privilège.

Je ne pourrais terminer ces remerciements sans mentionner ceux qui ont participé de plus près à ce mémoire ; ma deuxième famille, mes colocataires, Kevin, Harmony, Suzie et Thomas. Merci d'avoir endossé mes états d'âme, d'avoir été à l'écoute de mes réflexions les plus profondes, de m'avoir donné la confiance qui me manquait et d'avoir su en un rire me faire oublier ce pour quoi j'étais là. Les instants partagés avec vous constituent de loin la source de mon épanouissement.

La remise de ce mémoire coïncidant avec la fin de mon expérience au Québec, je souhaite remercier du fond du cœur tous ceux qui ont marqué ces cinq dernières années et qui ont fait en sorte que je tombe amoureuse de cette ville, Montréal. Bien évidemment, merci à toi Camille de m'avoir fait sentir, ici, chez moi.

# Introduction

«Ainsi, au projet d'une connaissance positive de la réalité se substitue le programme d'une maîtrise directe de tous les effets que nous pouvons « artificiellement » créer ou produire de manière méthodique et calculable dans tous les environnements spécifiés, de telle sorte que le « monde » n'est plus la totalité de ce qui est, mais l'ensemble de tout ce qu'on peut faire, prévoir, contrôler, transformer à volonté dans n'importe quel environnement : un monde entièrement centré sur « notre » puissance d'agir et qui en résulte directement, actuellement et surtout virtuellement, et ceci précisément parce que cette puissance nous échappe dans son objectivation, et qu'elle prend ainsi valeur de réalité première. » (Freitag, 2002 : 389)

Michel Freitag

Source d'espoir pour certains et d'apurement pour d'autres, l'Intelligence Artificielle (IA) suscite la fascination de beaucoup et éveille l'intérêt de tous. L'abondance des discours médiatiques sur l'IA démontre parfaitement cet engouement. La littérature de vulgarisation en IA se concentre exclusivement sur les enjeux, le fonctionnement, les apports bénéfiques et les risques potentiels. Avant d'entreprendre une action politique afin de minimiser les risques et maximiser les bénéfices de cette technoscience, il faut comprendre les enjeux qu'elle implique. Or, la compréhension de ces enjeux ne peut être effectuée sans l'examen approfondi de l'IA, pas seulement en tant que domaine de recherche, mais en tant que projet idéologique ; celui de transmettre l'intelligence à la machine. Dans cette optique, il est primordial d'établir sur quelle conception de l'être humain et de son intelligence s'est développé ce projet. Qu'est-ce que l'Intelligence Artificielle (IA) nous apprend de notre représentation de l'intelligence humaine? Quelle est la conception de l'intelligence des chercheurs et autres acteurs de l'IA pour que le projet d'une reproduction artificielle de cette qualité reconnue intrinsèque à l'humain soit concevable ? Ce présent mémoire se concentrera sur l'examen de ces deux principaux questionnements.

L'objectif de notre recherche sera de soulever les représentations sociales de l'intelligence en IA à travers les discours des chercheurs et autres acteurs du domaine dans une approche constructiviste. En envisageant l'intelligence comme un concept construit, nous verrons qu'aucune définition lui étant rattachée n'est univoque puisque toutes dépendent des représentations spécifiques à chaque société résultant des valeurs mises en avant au sein de celle-ci. Les représentations sociales opèrent sans que l'on ait toujours conscience qu'elles

existent ; c'est pourquoi il est fondamental de les clarifier. Soulever cette définition dans une perspective constructiviste permet non seulement de mieux appréhender le phénomène social qu'est l'IA, mais par-dessus tout, elle nous éclaire sur la vision de l'être humain qu'elle suppose. Nous tenterons de démontrer en quoi l'IA s'est développée sur la base d'une représentation réductrice de l'humain et de son intelligence sans jamais que les fondements de cette assise ne soient évoqués tant ils sont dissimulés aux confins des représentations sociales. D'autant que ce « paupérisme épistémique » (Robillard, 2019 : 3) n'est pas sans effets : de réels choix scientifiques en découlent entraînant la mise sur le marché d'applications concrètes qui guident et transforment notre quotidien. C'est pourquoi une réflexion de fond est de rigueur ; sur les assises de la discipline, mais aussi sur les notions qui l'identifient et les concepts qu'elle implique.

Dans l'univers des sociétés occidentales contemporaines, le progrès scientifique et technologique est assimilé à une évolution naturelle inéluctable. Pour preuve, l'objet technique est souvent perçu comme une extension de notre corps (Leroi-Gourhan, 1998 : 42). Or, nous tenterons de démontrer par l'entremise de cette recherche qu'il n'est que le produit construit de notre réalité sociale et dépend ainsi du contexte socioculturel spécifique dans lequel il s'implante. L'idéologie technoscientifique est entretenue par un modèle occidental puissant parce que dominant qui ne permet pas d'entrevoir au-delà de cette configuration épistémique d'une évolution adaptative de la technique. À la fois projet idéologique et objet technique, l'IA s'inscrit dans cette promesse de linéarité infinie de l'évolution technoscientifique. Sans prise de conscience de la convergence des valeurs entretenues par cette idéologie technoscientifique, l'analyse objective du phénomène est inconcevable. À partir de là seulement, il est possible de se détacher de l'univers épistémique dans lequel s'implante l'IA afin d'effectuer une analyse réflexive. Autrement dit, il s'agit de sortir du modèle dominant un instant pour enfin l'entrevoir ; sans quoi, on ne ferait que participer au maintien de cette idéologie. L'examen de la représentation de l'intelligence en IA passe donc nécessairement, selon nous, par une théorie critique du paradigme informationnel (que d'autres penseurs et théoriciens critiques de la modernité appellent « société postmoderne » (Lyotard, 1979 : 11). Sans « la reconnaissance du rôle fondateur de la cybernétique » (Lafontaine, 2004 : 18) dans notre société, le débat sur l'IA est incomplet et incompris. Unique horizon paradigmatique, c'est à travers le modèle informationnel qu'a émergé le projet d'une IA et c'est à travers la compréhension de celui-ci seulement, que le débat concernant les enjeux gigantesques que cette technoscience provoque, peuvent être, selon nous, discutés. Un regard réflexif et critique envers le paradigme informationnel n'est pas négligeable afin de saisir l'ampleur du problème philosophique qu'est

celui de « l'attribution du prédicat d'intelligence à un système artificiel », pour reprendre les termes du philosophe Jean Robillard (Robillard, 2019 : 12). La représentation de l'intelligence en IA est calibrée sur ce même paradigme. Loin d'être la chasse gardée des chercheurs et transhumanistes, cette conception de l'intelligence est partagée instinctivement par le plus grand nombre ; sans quoi elle serait facilement contestable. L'IA n'est d'ailleurs concevable que par le biais de cette conception commune. Les fondements de l'IA reposent sur une vision du monde omnisciente dont la cybernétique est le tributaire ; d'où l'appellation de « paradigme ». C'est aussi pourquoi elle confère autant de considération de part et d'autre des secteurs de l'économie.

Ce mémoire entend rendre compte des intérêts sociaux affectés par la rationalité instrumentale technocratique. Parce que la technoscience est centrale dans les sociétés occidentales postmodernes et que l'IA est un des domaines sur lesquels on mise le plus aujourd'hui, elle constitue le reflet de nos sociétés. En étudiant l'intelligence à travers cette technoscience, on étudie donc l'intelligence à travers les représentations des sociétés occidentales contemporaines. Autrement dit, concentrer notre attention sur cet objet de recherche précis (l'IA), permet de soulever une réalité sociale plus large. L'IA, en tant que technoscience, bouscule les assises sociales. Elle altère le rapport au monde et à soi-même. À ses prémices, l'IA correspond à la consécration technologique du fantasme ancien de reproduction artificielle de ce qui définit le mieux l'Homme et le distingue des autres espèces. Or, c'est à l'intelligence qu'est attribuée cette propriété. En tant qu'objet de recherche, l'intelligence permet donc d'entrevoir les frontières tangibles de l'être humain dans les sociétés technocratiques occidentales. En plus de rendre compte de l'altération causée par les nouvelles technosciences sur le rapport de l'individu à lui-même et au monde qui l'entoure, l'étude de l'IA constitue un point de départ pertinent à l'étude des représentations sociales de l'être humain.

Analyser les représentations de l'intelligence en IA impose un retour sociohistorique sur l'évolution du projet idéologique puis disciplinaire. Nous retracerons donc dans un premier chapitre l'histoire dans laquelle s'inscrit l'IA. La perspective sociale ajoutée à l'approche historique permet d'entrevoir toute l'ampleur du phénomène. Nous ne nous limiterons donc pas à décrire l'évolution de la discipline de ses prémices à aujourd'hui. L'IA fait partie des nombreuses créatures artificielles façonnées à « l'image de l'Homme » (Breton, 1995) dont la tradition est ancestrale. Le survol de cette genèse nous permettra de comprendre en quoi l'IA a

été pensé à partir d'assomptions attribuées à l'être humain et à son intelligence. Cette représentation spécifique évolue dépendamment de l'univers épistémologique dans lequel nous nous trouvons. Nous verrons donc dans ce même chapitre en quoi l'IA n'aurait pu être pensée en dehors du modèle informationnel, un paradigme introduit par la cybernétique. De la légitimation d'un fantasme par le fondement d'une discipline de recherche scientifique, à sa concrétisation technologique dont les fondements idéologiques sont peu à peu oubliés ; nous rentrerons à ce terme dans l'évolution historique de l'IA.

Le deuxième chapitre sera consacré aux définitions de l'intelligence. Nous chercherons à établir sur la base de la multiplicité et diversité de définitions assignées à cette notion que l'intelligence constitue un concept construit. Il s'agira de démontrer que toute tentative de définition de l'intelligence dépend de l'univers épistémologique dans lequel il s'implante et n'est donc « exacte » qu'au sein de cette configuration. Le concept d'intelligence tel que nous le connaissons aujourd'hui s'est construit au fil des grandes approches théoriques de notre civilisation. L'idée même d'une reproduction artificielle concrète au sein d'un support matériel n'aurait pu être pensable sans cette exacte même évolution du concept. Nous étudierons donc l'intelligence en tant que concept construit à la lumière des représentations sociales dans un premier temps, et au fil des grandes approches historiques dans un second temps. À partir de ce constat, il sera possible d'analyser les représentations sociales de l'intelligence en IA. De la rétroaction à l'optimisation en passant par la résolution de problème, la définition de l'intelligence en IA est en constante évolution avec son temps. Nous verrons dans cette partie en quoi l'éloignement progressif de l'ambition idéologique de base vers une simple optimisation des capacités de l'humain et de la machine retranscrit un réductionnisme de plus en plus profond dans la représentation de l'Homme et de son intelligence. Nous montrerons pour finir que l'IA n'est envisageable que par la déconstruction du dualisme nature / technique qui, non seulement, légitime le projet de rendre la machine intelligente, mais justifie la possibilité d'une telle démarche. Le courant de pensée transhumaniste illustre parfaitement cette indifférenciation ontologique introduite par le brouillage des frontières originaires calibré par les catégories dualistes. Nous verrons notamment que l'idéologie transhumaniste repose autant sur l'artificialisation de l'intelligence humaine que sur la naturalisation de l'intelligence artificielle.

Le descriptif de la démarche méthodologique emprunté pour ce mémoire sera effectué au troisième chapitre. Après avoir établi notre objet de recherche, et explicité la méthode de l'analyse de discours, nous procéderons à la construction des données sociologiques. Il s'agira

de présenter les catégories d'analyses, les matériaux analysés ainsi que les types d'acteurs des discours sélectionnés.

Le quatrième et dernier chapitre constituera la partie empirique de ce mémoire : l'analyse de discours. Aucune méthode scientifique ne met plus en lumière les représentations sociales que l'analyse discursive. Cette méthode qualitative permettra en effet de soulever les conceptions des acteurs de l'IA à travers les discours de vulgarisation. Il s'agira donc d'interpréter les propos des chercheurs, transhumanistes, entreprises, gouvernements et journalistes autour de l'IA. L'incorporation du transhumaniste à notre recherche est pertinente à la compréhension des discours en IA. Le maintien d'une ambition idéologique chez les transhumanistes oblige une formulation théorique difficilement repérable dans les propos des chercheurs contemporains en IA. Cette forme de discours dévoile plus explicitement les représentations sociales de l'intelligence en IA et sert ainsi l'élaboration d'une réponse plus approfondie à notre problématique de recherche. La codification établie au préalable, via le logiciel Atlas ti, nous permettra d'établir des catégories d'analyse, lesquelles seront développées au cours de ce chapitre quatre. Ayant fait le choix de ne sélectionner que des discours récents, l'analyse se concentrera sur la représentation contemporaine de l'intelligence en IA : soit le critère de l'optimisation. Nous mettrons ensuite en perspective le réductionnisme de cette représentation de l'Homme et de son intelligence, pour enfin mettre en relief l'ambivalence de l'IA entre nature et technique.

# Chapitre I.

## Retour sociohistorique de l'Intelligence Artificielle

À la fois produit social et productrice de nouvelles réalités sociales, l'Intelligence Artificielle (IA) participe grandement au bouleversement que les technosciences entraînent de nos jours. Afin d'appréhender la représentation de l'intelligence en IA, une analyse sociohistorique du phénomène est de vigueur, ce à quoi s'attardera ce premier chapitre. La définition de l'intelligence établie en IA (celle qui permet de penser que celle-ci peut être reproduite au sein d'une machine) dépend très largement de la représentation de l'être humain établi dans un univers épistémologique bien spécifique. « Le récit fondateur de l'IA est d'abord celui de la construction d'une représentation de l'humain. » (Breton, 1995 : 103) L'IA, avant de devenir une discipline de recherche, constitue avant tout un projet idéologique. Les fondements de cette idéologie se sont construits autour d'un imaginaire traversé par une longue tradition occidentale. Au même titre que la statue de Pygmalion, le Golem du Rabbin Loewl ou même l'ordinateur, l'IA illustre une des nombreuses représentations de l'être humain. Ainsi, nous démontrerons dans une première partie en quoi le projet de créer une Intelligence Artificielle se rattache à la genèse des créatures artificielles façonnées à l'image de l'Homme, pour reprendre l'expression de Philippe Breton. Il conviendra ensuite de se demander, comment et au sein de quel univers épistémologique le projet de l'IA est-il pensable. Nous verrons en quoi la définition de l'intelligence telle qu'elle est comprise en IA prend forme au sein du modèle informationnel, un paradigme introduit par la cybernétique. Dans une seconde partie, nous étudierons la transition de l'IA en tant que projet idéologique, à l'IA en tant que discipline de recherche scientifique. Il s'agira d'analyser les métaphores de l'esprit humain (neuronales et computationnelles) sur lesquelles s'appuie l'IA afin de légitimer la possibilité de concrétisation scientifique d'un projet qui n'était alors qu'un fantasme. Après un survol des prémices de l'IA, nous rendrons compte de la légitimation d'un projet qui n'était alors qu'un fantasme à travers l'examen de l'IA symbolique, premier courant avéré d'une discipline scientifique naissante. Du béhaviorisme pour les partisans de l'analogie neuronale au cognitivisme pour les disciples de l'analogie computationnelle, nous verrons que les programmes d'IA sont construits dans la mouvance des grandes approches scientifiques en vigueur. Nous analyserons finalement la concrétisation de l'IA en tant que véritable technoscience. Nous retracerons dans cette optique les méthodes (système expert), concepts

(apprentissage) et approches (connexionnisme) qui expliquent l'envolée technologique que connaît l'IA de nos jours.

## **1. L'IA : une créature artificielle ...**

### **1.1. ... à l'image de l'Homme**

Le retour sociohistorique de l'IA ne se limitera pas au premier ordinateur créé, ni même à la première fois où le terme « intelligence artificielle » a été employé, puisque le projet d'une IA s'inscrit dans la continuité de la genèse des créatures artificielles façonnées à l'image de l'Homme. Cette genèse, Philippe Breton la retranscrit dans l'ouvrage *A l'image de l'Homme. Du Golem aux créatures virtuelles* (Breton, 1995). Toutes ces créatures divergent dans leurs formes et caractéristiques, car les cadres de représentation dans lesquels elles ont été produites sont profondément différents. En revanche, la logique est la même. C'est cet ordre de signification commun qui compose la continuité sur laquelle s'est établie l'IA. L'IA n'est autre qu'une créature artificielle façonnée par l'Homme à l'image de l'Homme. Dans le processus de création, l'être humain fait appel à un imaginaire ancré dans un contexte spécifique. La continuité de ces récits mettant en lumière les figures artificielles démontre parfaitement en quoi l'étude de ces créatures virtuelles est un point de départ pertinent à l'étude de l'être humain : « miroir essentiel dans lequel l'Homme est confronté à sa propre image » (Breton, 1995 : 7). L'attention se porte sur ce qui est considéré comme caractérisant le mieux l'humanité. De la beauté de la femme à travers la statue Galatée, à « l'intelligence » de l'ordinateur, nous verrons en quoi ces qualités jugées essentielles en leur temps sont la démonstration de la représentation de l'Homme par l'Homme.

#### *Pygmalion*

Principalement rapporté par Ovide dans les *Métamorphoses*, le mythe grec de Pygmalion illustre parfaitement la fondation de l'IA établie au sein de la tradition ancestrale de réplique de l'être humain. Le sculpteur Pygmalion, par dégoût de la femme en chair et en os, façonne une statue d'ivoire à l'image de la femme idéale qu'il nomme Galatée. Le sculpteur tombe amoureux de sa création artificielle, si bien qu'Aphrodite décide de lui donner la vie (Ovide, 1966). Dans l'imaginaire des créatures artificielles, l'Homme est créateur de sa propre image. Ainsi, Pygmalion choisit les modalités de création selon ses propres critères dépendants d'un univers spécifique. Ne trouvant pas son compte parmi les êtres de nature, il est néanmoins

capable de se représenter ce qui pour lui constitue l'être féminin idéal. Avant de prendre vie, la statue n'est ni douée de parole ni de raison. Le sculpteur est amoureux de son apparence, de la beauté physique. Le mythe de création d'une femme artificielle par un Homme mortel nous expose à la représentation de l'être humain dans le contexte de la Grèce Antique. La matérialité du corps, partie intégrante et indissociable de l'être humain, détient au sein de cette représentation une place prépondérante. Le sculpteur ne se contente pas de créer la femme telle qu'il se la représente, il façonne la femme « idéale ». L'idéalité du corps au sein de cette représentation est identifiée par la beauté (selon les normes de l'époque). Autrement dit, la représentation de l'être humain se fait à travers son corps, et c'est la beauté de ce corps qui semble être la caractéristique le définissant le mieux, représentant ainsi la qualité jugée la plus valorisée de l'humain en ce temps.

### *Le Golem*

Dans la continuité des créatures artificielles façonnées par l'Homme à l'image de l'Homme, la figure du Golem est certainement celle qui illustre le mieux la tradition ancestrale d'autoreproduction dont provient l'IA. C'est à la mythologie juive que l'on doit l'introduction de cet humanoïde fait d'argile créé dans l'optique d'assister et défendre son maître. Dénué de parole et de libre arbitre, le Golem a la capacité d'agir et réagir en fonction de ce qui lui a été prescrit. Déjà mentionné dans le *Talmud* (l'un des textes fondamentaux du judaïsme), c'est la tradition populaire juive qui attribua au XVIIe siècle, la création du Golem au Rabbin Loewl (Wiener, Wall-Romana, Mopsik et Scholem, 2001 : 113). C'est un être humain qui façonne la créature de ses mains selon l'image qu'il se fait de sa propre espèce. Le Golem prit vie lorsque le rabbin inscrit sur son front les lettres « EMET », soit un des noms donnés à Dieu dans le judaïsme et signifiant aussi « vérité ». Lorsqu'un jour le rabbin perd le contrôle de sa créature, et qu'il n'a d'autres choix que de lui retirer la vie, il lui suffit d'effacer de son front la première lettre de « EMET », ce qui donne « MET » et qui signifie « mort ». En retirant « la vérité » et le nom de Dieu, on lui retire sa faculté de raisonner et instantanément, le corps de la créature redevient glaise (McCorduck, 2004 : 14). Alors que dans le mythe de Pygmalion le corps et l'esprit prennent vie indépendamment l'un de l'autre, la vie du Golem dépend essentiellement de sa faculté de raison. La matière du corps du Golem est nécessaire, mais pas primordiale à son esprit, il sert simplement de support. Cette légende nous éclaire sur la représentation de l'être humain dans le contexte du XVIIe siècle : l'Homme n'est plus seulement corps et matière, il est un être de raison. Le concept de raison est déjà présent dans la pensée grecque antique sous le terme *logos*. La nuance se trouve dans l'entendement du concept : tandis que *logos*

désigne une forme de pensée caractérisée par l'habilité à employer une langue, chez Descartes, l'utilisation de la langue constitue seulement un outil permettant à l'Homme d'émettre un jugement de valeur, soit de raisonner. C'est cet attribut, à travers le mot « Emet » (vérité), qui tient le Golem en vie et lui donne la capacité de suivre des ordres. La vérité est associée à la capacité de juger, c'est-à-dire distinguer le vrai du faux et prendre position en conséquence. C'est pourquoi la représentation de l'humain se réfléchit dans cette légende à travers le paradigme cartésien.

En effet, Descartes introduit un nouveau modèle de pensée à travers lequel notre rapport au monde se transforme. C'est désormais l'intelligence qui est jugée composante essentielle de l'être humain. Au sein du paradigme cartésien, l'intelligence est caractérisée par la raison ou ce que Descartes appelle aussi « le bon sens ».

« Cela témoigne que la puissance de bien juger et distinguer le vrai d'avec le faux, qui est proprement ce qu'on nomme le bon sens ou la raison, est naturellement égale en tous les hommes ; et ainsi que la diversité de nos opinions ne vient pas de ce que les uns sont plus raisonnables que les autres, mais seulement de ce que nous conduisons nos pensées par diverses voies, et ne considérons pas les mêmes choses. » (Descartes, 1987 : 112)

Descartes dévoile à travers cet extrait du *Discours de la méthode* sa définition de la raison. Ce serait la faculté à juger et discerner ce qui est avéré, fondé et juste de ce qui ne l'est pas. L'intelligence de l'humain ne tient donc pas du raisonnement tel qu'on l'entend aujourd'hui, mais du jugement, concept proprement philosophique (Cassan, 2007 : 133). Dans ce paradigme, l'intelligence est donc caractérisée par la faculté de jugement de l'Homme; elle serait égale à tous les êtres humains et ce serait précisément ce qui nous distinguerait des bêtes. Un jugement que Descartes croit impartial, puisque réfléchi, donc vrai. Nous reviendrons sur ce point au chapitre 2. L'intelligence, telle qu'elle est comprise au sein du paradigme cartésien (Perler, 2004 : 463), devient alors le caractère le plus constitutif de l'humain.

Avec le temps, on verra le Golem davantage comme un automate que comme une créature divine. Encore là, on y voit le reflet de la représentation de l'Homme dans la société occidentale qui évolue avec le développement de la science moderne. « L'Homme dans cette perspective est toujours un être créé, mais la porte est ouverte à ce qu'il soit une créature de la science plutôt qu'une créature divine. » (Breton, 1995 : 94) C'est cette perspective qui permit à Wiener, père de la cybernétique, de s'approprier la légende comme source d'inspiration de ses travaux. « La machine, comme je l'ai indiqué, est l'homologue moderne du Golem du Rabbine de Prague ». (Wiener, 2001 : 111) Wiener prétendait même être un des descendants du Rabbine Loewl. En développant le modèle de la cybernétique, il est à l'origine d'un nouveau paradigme redéfinissant les frontières de l'humain et au sein duquel le projet de l'IA s'est

développé. C'est aussi pourquoi la légende du Golem est un récit essentiel à la compréhension de l'avènement de l'IA.

### *L'ordinateur*

Une autre créature fondamentale dans la lignée de l'IA est bien évidemment l'ordinateur. L'EDVAC, considéré comme le tout premier, fut créé en 1945 par le mathématicien et physicien John Von Neumann (Neumann, 1945). L'ordinateur est né dans l'optique de création d'un cerveau artificiel à l'image du cerveau humain (Breton, 1995 : p.14). Le raisonnement humain se limite selon lui au résultat d'un traitement de l'information neuronale (Breton et Proulx, 2016 : 86). La démarche de production du cerveau artificiel se trouverait alors dans la compréhension des modalités de ce traitement. La réplique de l'intelligence humaine, dans cette perspective, serait alors accessible à la science. L'ordinateur constitue donc la première étape vers un objectif plus ambitieux de reproduction artificielle de l'être humain, ou sinon de création d'une intelligence artificielle. Philippe Breton nous invite à voir dans les plans de l'ordinateur « le reflet d'une représentation particulière de l'humain » (Breton, 1995 : p.138). On considère à l'époque l'ordinateur comme une « machine à exécuter les tâches de la pensée » (Breton et Proulx, 2016 : 303). Or, ses aptitudes se limitent à la gestion, au calcul et au traitement de textes et d'images. Cette réduction en dit long sur la représentation de l'humain au sein du modèle informationnel dont nous analyserons les fondements prochainement. On ne parle plus de faculté de jugement et de discernement, mais de traitement d'informations et de résolution de problèmes. C'est toujours l'intelligence qui est jugée être l'essence de l'être humain et la qualité la plus démonstrative de notre humanité, mais celle-ci est désormais considéré comme instrumentale et opératoire. L'intelligence n'est plus l'affaire de la philosophie, mais des sciences naturelles et des mathématiques.

Que ce soit au travers de la mythologie, de la littérature, de l'automatisation ou de l'informatique, nous avons vu que Philippe Breton nous invite à voir en les créatures artificielles le reflet ponctué des représentations de l'être humain. Le regard porté sur ces créatures est en fait un regard sur l'être humain. Pygmalion, le Golem, l'ordinateur et l'IA ont pour dénominateur commun d'être créés par l'Homme à l'image de l'Homme. Chacune de ces figures nous informe sur la représentation de l'être humain au sein de l'univers épistémologique dans lequel il s'implante. On l'a vu, ces représentations diffèrent. Alors qu'au travers du mythe grec de Pygmalion on voit en la beauté la qualité définissant le mieux l'humanité, c'est l'intelligence dans le mythe du Golem qui prend sa place et qui s'explique à travers le

paradigme cartésien. Le regard porté sur cette qualité intrinsèque à l'être humain est encore philosophique. Avec la création de l'ordinateur en revanche, la représentation de l'intelligence devient mathématique et opératoire.

Nous verrons dans cette prochaine partie comment l'intelligence, devenue centrale à partir de Descartes, est devenue information dans le paradigme des sociétés occidentales contemporaines dans lequel s'est construite l'IA communément appelée « modèle informationnel ».

## **1.2. ... au sein du modèle informationnel.**

L'IA est le produit d'une recherche de similitude qui passe par la compréhension du modèle humain. Cette compréhension prend racine au sein d'un cadre épistémologique bien spécifique : le modèle informationnel, dont il convient désormais d'étudier les fondements.

### *Théorie de l'information*

À la base du modèle informationnel se trouve la théorie de l'information. Claude Shannon, mathématicien renommé, élabore à travers cette théorie mathématique un modèle qui permet de décrire certains aspects du réel. Le paradigme informationnel constitue, lui, une tentative de réduction du réel à ce modèle (Breton et Proulx, 2016 : 131) . La « théorie mathématique de la communication de l'information », nom donné à la théorie de l'information par son auteur, est la formule instrumentale de l'information sur laquelle repose tout un paradigme. Elle est donc à la base d'une rupture épistémologique fondamentale.

Shannon, alors ingénieur pour BELL, est chargé de rendre compte des conditions de transports, de codages et de dégradation du signal dans le domaine de la téléphonie. L'objectif de la compagnie est de transmettre les messages de la façon la plus économique et la plus fiable possible. C'est dans ce contexte qu'il développe la théorie de l'information. Cette théorie probabiliste permet de quantifier le contenu de l'information dont le langage binaire (à la base du codage informatique) satisfait une distribution statistique assez précise pour réduire l'incertitude à la transmission d'un message. Autrement dit, l'objectif est de quantifier le message transmis de l'émetteur au récepteur. Shannon expose clairement dans l'article «A Mathematical Theory of Communication» la volonté de laisser de côté les propriétés sémantiques et physiques du message pour s'intéresser exclusivement aux propriétés mathématiques et communicationnelles. « Frequently the messages have meaning ; that is they refer to or are correlated according to some system with certain physical or conceptual entities.

These semantic aspects of communication are irrelevant to the engineering problem. » (Shannon, 2001 : 3) Il ne dit pas que la signification des messages transmis n'est pas importante en soi, seulement que celle-ci appartient au domaine de l'interprétation humaine et n'entre en aucun cas dans l'exercice de cette mesure de l'information.

### *La cybernétique*

C'est la cybernétique de Wiener qui constituera une réelle révolution épistémologique en proposant un modèle englobant toutes les théories à partir de cette théorie de l'information développée par Shannon. À la différence de cette dernière qui cherche à expliquer certains concepts du réel par le biais de la formulation mathématique, nous verrons que la cybernétique développée tend à réduire le réel au modèle informationnel.

La cybernétique est un terme proposé par le mathématicien Norbert Wiener en 1948. Il l'a défini comme étant une « théorie entière du contrôle et de la communication, aussi bien chez l'animal que dans la machine » (Wiener, Le Roux, Vallée et Vallée, 2014 : 70). Céline Lafontaine dans l'article « The Cybernetic Matrix of French Theory », nous rappelle que la cybernétique a été pensée dans un contexte particulier indissociable de sa nature et de son objectif. Au sortir de la Seconde Guerre mondiale, un profond pessimisme politique se fait ressentir résultant de la défaite de l'idéal humaniste avec l'holocauste et de la menace du communisme à venir. Ce pessimisme politique aurait accompagné un fort optimisme pour les technosciences résultant en la création de la cybernétique. Si l'on regarde sa racine étymologique, le mot « cybernétique » fait référence à l'art de piloter, et par extension à « l'art de gouverner les hommes » (Lafontaine, 2004 : 40). Pour Lafontaine, on y voit là l'ambition de créer une machine capable de contrôler, de gouverner les Hommes et les sociétés et de pallier aux faiblesses de l'Homme, qui clairement, a échoué. Il s'agit donc de débarrasser la raison de l'humain pour la transmettre à la machine, car, selon Wiener et ses compères, ce n'est pas la raison, mais l'humain qui est faillible.

Les trois concepts fondamentaux de la cybernétique se comprennent selon ce contexte sociohistorique et l'état d'esprit dans lequel se trouvait Norbert Wiener. Pour paraphraser Philippe Breton, dont sa définition est retranscrite dans *L'Empire Cybernétique* de Céline Lafontaine, *l'entropie*, seconde loi de la thermodynamique, fait référence à tout système isolé et fermé qui tend vers un état de désordre maximal (Lafontaine, 2004 : p.41). Le concept d'entropie dans la cybernétique est central et relatif au pessimisme politique de Wiener : ce n'est plus l'Homme qui est désigné comme hostile à la société, mais l'entité du désordre, de la désorganisation. C'est donc *l'entropie* qu'il s'agit de combattre afin de sauver l'humanité de la

détérioration. « De même que l'entropie est une mesure de désorganisation, l'information fournie par une série de messages est une mesure d'organisation. » (Wiener, 1962 : 207) Ce serait donc dans *l'information* (autre concept phare de la cybernétique) que se trouverait le moyen de combattre l'entropie ou du moins d'en ralentir le processus. C'est ce concept, dont la définition provient très largement de la théorie de l'information de Shannon, qui permet à Wiener d'adopter une position déterministe optimiste face au progrès qui consiste, selon lui, en « l'amélioration du contrôle et du traitement de l'information » (Lafontaine, 2004 : 43). La théorie de l'information permet de codifier, d'améliorer et de maîtriser à première vue la machine, mais aussi le vivant. Le dernier concept central de la cybernétique est celui de *la rétroaction*. Il se rapporte à la conception circulaire et infinie de la communication qu'avait Wiener. *La rétroaction* est le processus par lequel l'information est identifiée, appropriée puis appliquée dans l'optique de guider et contrôler l'action. Nous verrons en quoi d'un point de vue informationnel la définition de la rétroaction se trouve être la même que celle de l'intelligence pour les tout premiers concepteurs des machines d'IA au chapitre II. En résumé, l'un des buts premiers de la cybernétique est de combattre l'entropie et c'est l'information qui a le pouvoir de s'y attaquer via la rétroaction. On comprend donc selon quelle assise le modèle informationnel est celui à partir duquel l'entièreté du monde peut s'expliquer selon Wiener et ses contemporains.

La cybernétique en tant que modèle idéologique et théorique prétend donc pouvoir englober tous les aspects du monde et en expliquer tous les phénomènes. La volonté de base est d'ailleurs de promouvoir une conception qui englobe les théories, concepts et applications des domaines naissants tels que la théorie de l'information, l'informatique, l'électronique, l'automatique, les sciences cognitives, etc. (Wiener, Le Roux, Vallée et Vallée, 2014 : 70). La cybernétique constitue en ce sens un modèle omniscient qui tient son caractère universel à un projet interdisciplinaire, autrement dit, à un projet dont la convergence des sciences et disciplines adjacentes est le principe fondateur de l'entreprise. Au sortir de la guerre, dès 1946, des rencontres rassemblant des scientifiques renommés dans leur champ disciplinaire respectif, communément appelé « conférences Macy », se tiendront à dix reprises dans le but de discuter d'une possible unification sous un même modèle des différentes approches théoriques et méthodes. « Même si elle a pu donner lieu à des confusions conceptuelles, l'interdisciplinarité des conférences Macy place d'emblée la cybernétique sous le signe d'une nouvelle Renaissance rendant enfin possible l'unification des savoirs. » (Lafontaine, 2004 : 40)

L'ampleur de la cybernétique ne s'arrête donc pas à la simple ambition de concevoir une nouvelle théorie scientifique englobant toutes les connaissances, elle représente aussi une

« révolution épistémologique » (Lafontaine, 2007 : 29), (Josset, 2016 :9). Un paradigme nouveau est instauré rompant définitivement avec l'ancien paradigme cartésien (Josset, 2016 :10). Ce dernier paradigme s'appuyait sur une méthode analytique linéaire de décomposition des éléments de la nature afin de comprendre le fonctionnement du monde, ce que Norbert Wiener appelle « méthode fonctionnelle » (Breton, 1995 : p.108). C'est sur cette méthode d'approche du réel introduite par Descartes que se fondent les sciences classiques. Si l'objectif de conquête de la nature est le même au sein des deux paradigmes (cartésien et cybernétique), la méthode diffère. Philippe Breton, dans son ouvrage explique ce qui fait la spécificité de la méthode d'approche du réel introduite à travers la cybernétique que Norbert Wiener nomme la méthode comportementale (en opposition à la méthode classique fonctionnelle) : c'est la relation des objets entre eux et leurs interactions avec l'environnement sur laquelle on devrait focaliser notre analyse et non pas sur la propriété spécifique de chacun de ces objets.

Suivant cette logique analysée par Philippe Breton, on comprend désormais pourquoi le père de la cybernétique attache une importance première à la dimension informationnelle des éléments ; la dimension matérielle n'étant que secondaire (Breton, 1995 : 138). C'est la transmission de l'information (l'interaction) qu'il pense essentielle à considérer et non pas le support matériel servant à la transmission (la propriété de l'objet). Tout s'explique désormais selon cette approche informationnelle. « Ce modèle purement différentiel conduit à l'annulation ontologique de toute distance entre vivant et non-vivant. La cybernétique subordonne en fait la vie au principe informationnel. » (Lafontaine, 2004 : 43) En ce sens, le corps de l'humain n'étant que support matériel, l'être humain est ainsi réduit à un objet informationnel au même titre que la machine. L'indifférenciation ontologique de l'homme et de la machine est ainsi justifiée et rompt définitivement avec le dualisme cartésien qui prônait une séparation de l'esprit (du côté du divin) et du corps (du côté de la machine). Le père de la cybernétique, lui, place l'esprit du côté de la machine tandis que le corps n'est plus que support matériel non pertinent.

Le modèle informationnel propose donc une vision du monde global dans laquelle l'information et les processus de communication jouent un rôle déterminant au point que tous les objets composant le monde sont comparables sous la même échelle : l'échelle informationnelle. Dans la perspective informationnelle, il n'existe plus de séparation ontologique entre l'humain et la machine. Plus qu'un simple modèle théorique, le modèle informationnel constitue donc un véritable paradigme au sein duquel la façon de penser la vie et de se penser soi-même en tant qu'être humain est profondément bouleversée. La représentation de l'Homme en tant qu'information dans ce cadre épistémologique légitime

d'ailleurs la reproduction de l'être humain et la permet. Autrement dit, c'est cette représentation de l'humain dans cet univers épistémologique spécifique qui a ouvert la possibilité d'entrevoir une reproduction concrète de l'intelligence humaine par des moyens scientifiques, soit l'IA.

Dans la compréhension du modèle informationnel se trouvent les assises de l'explication du passage de l'Intelligence Artificielle en tant que projet idéologique de création d'une « machine pensante » au domaine de recherche scientifique qu'elle constitue à ce jour. Il convient dorénavant d'analyser les composants et événements majeurs qui constituent cette transition.

## **2. L'IA : du projet idéologique à la discipline scientifique**

### **2.1. Légitimation d'un fantasme.**

Bien que l'IA s'incarne entièrement dans le modèle informationnel, ces fondations vont au-delà des méthodes théoriques établies par la cybernétique.

#### 2.1.1. Les prémices

##### *Turing*

Alan Turing fut un des premiers mathématiciens à tenter de décrypter le fonctionnement de la pensée humaine en cherchant à la répliquer au sein d'une machine qu'il appelle « machine universelle » (Turing, 1937 : 241). Ces travaux et découvertes, indissociables du contexte militaire, ont largement contribué au développement de l'ordinateur en inspirant Neumann dans l'élaboration des plans de l'EDVAC. C'est en raison de ce vif intérêt pour l'esprit humain qu'Alan Turing est l'un des premiers à débattre de la possibilité de développement d'une IA. En 1950, le célèbre mathématicien publia un article qui compléta sa renommée : « Computing Machinery and Intelligence » dans le magazine anglais *Mind*. Dans cet article, il pose la question : « Can a machine think? » (Turing, 1950 : 433) Pour l'auteur, si une machine, certes, ne peut pas penser au même titre qu'un être humain doué de raison et d'intentionnalité, elle peut toutefois développer la capacité de « penser » d'une autre manière, non moins importante et de manière plus performante que l'être humain. C'est le premier à postuler que la machine puisse faire autant, sinon plus, que les êtres pensants que nous sommes. Dans cet article, il tente d'expliquer comment. Pour le mathématicien, le cerveau n'est rien de plus qu'une « machine

inconsciente qui produit des réactions lorsqu'elle est stimulée » (Bouée et Roche, 2017 : 29). Par l'analogie machinique, il remet ainsi en question le concept même de conscience chez l'humain : la machine n'est pas dotée de conscience tout simplement parce que l'Homme non plus n'est pas nécessairement outillé de cette qualité qui lui est habituellement assignée. L'indifférenciation ontologique entre la machine et l'être humain est ici évidente et assumée. Il développe un test, aujourd'hui appelé « Test de Turing », dont le processus est retranscrit dans ce même article. Ce test permet de détecter le moment où la machine aura atteint le stade de l'intelligence : si un interlocuteur ne peut distinguer l'être humain de la machine en communiquant avec ces deux entités, alors, on pourra dire que la machine aurait atteint le même niveau d'intelligence que celle de l'humain. Ce test en dit long sur la représentation particulière de l'ingénieur sur l'intelligence, et plus largement de l'être humain. Le test de Turing témoigne parfaitement de la représentation de l'être humain comme être informationnel dont le corps ne fait office que de support négligeable. Cette perspective lui permet d'avancer qu'il est possible de créer une machine intelligente à l'image de l'Homme. Par-là, il ouvre la porte à la considération de développement de l'IA de manière concrète. Ainsi, par la légitimité accordée à ses travaux dans l'univers scientifique, il représente le premier médiateur de la transition de l'Intelligence Artificielle en tant que projet idéologique à l'IA en tant que projet de recherche scientifique.<sup>1</sup>

### *L'analogie neuronale*

Parallèlement aux publications et travaux de Turing, les recherches en neurosciences introduisent de nouvelles idées et modèles théoriques déterminants pour la recherche en IA en tant que domaine de recherche scientifique : l'étude de l'esprit est appréhendée sous la métaphore neuronale. En d'autres termes, l'esprit humain s'expliquerait par le fonctionnement du système neuronal. Cette conception donne lieu au courant connu aujourd'hui sous le nom de connexionnisme. Le terme n'apparaît qu'en 1980 pour contrer le courant computationnel (nous y reviendrons), mais ses assises sont fondées à la même période que la cybernétique. C'est donc en connivence avec le modèle cybernétique que les premiers modèles mathématiques et informatiques de la pensée humaine ont été développés avant même que le terme Intelligence Artificielle ne fasse son apparition au sein de la communauté scientifique. Le lien entre la

---

<sup>1</sup> L'hypothèse selon laquelle l'IA est encore de nos jours épris des fondements idéologiques sera posée lors de notre analyse de discours et la question de « transition » sera alors remise en question.

cybernétique et l'analogie neuronale de l'esprit se crée à partir de l'approche béhavioriste. Cette branche de la psychologie, née au début du 20<sup>e</sup> siècle se concentre sur le comportement de l'individu en interaction avec son environnement. Son étude s'appuie sur la mesure scientifique et l'expérimentation opérationnelle (Nilsson, 2010 : 20). Elle ne fait pas appel au psychisme et est mesurable objectivement. Watson, fondateur du courant, pose les bases du behaviorisme dans un article intitulé « Psychology as the Behaviorist views it »(Watson, 1913). Le béhaviorisme analyse l'association entre le stimulus en provenance de l'environnement et la réponse comportementale émise par l'organisme. L'environnement (stimuli) agit sur l'organisme (la boîte noire) qui réagit par un comportement (la réponse). Le béhaviorisme consiste en l'étude d'un système sans considérer son fonctionnement interne inaccessible à l'investigation scientifique puisque non observable à la différence des relations stimulus-réponse. Le terme de *boîte noire* pour représenter l'organisme, a été intégré au béhaviorisme (et non l'inverse) qu'à partir de la théorisation de Wiener dans l'essai *Cybernétique et société* en 1948 (Wiener, 1962). La cybernétique constitue en quelque sorte la modélisation de cette approche. Selon cette perspective, la *boîte noire* constitue l'organisme inexplorable de toute organisation. Elle doit être appréhendée sous l'angle de ses interactions en s'intéressant exclusivement à ses entrées et sorties et non à son mécanisme interne. Le simple fait de savoir que la *boîte noire* est l'endroit où la transformation des entrées en sorties s'opère serait suffisant à l'étude exclusive des échanges, ce qui serait suffisant pour comprendre le fonctionnement de toute entité. Nul besoin, dans cette perspective, de rentrer dans l'étude des mécanismes qui s'opèrent au sein même de cette *boîte noire*.

C'est la loi de Hebb formulée en 1949 par le neurophysiologiste Donald Hebb qui constitue, encore à ce jour, le fondement du courant connexionniste. Selon cette loi, les connexions synaptiques entre deux neurones se renforcent ou se créent sous l'effet de leur activation simultanée induite par l'environnement. Dans cette perspective, la cognition est parallèle (ensemble de neurones intégrant de l'information simultanément) et distribuée (connaissance localisée dans la multitude de connexions synaptiques qui relient les neurones entre eux). Hebb dit remettre en question l'approche béhaviorisme en voulant explorer les processus neuronaux se déroulant dans la *boîte noire B* (Hebb, 1974). Au vu de sa modélisation, son appréhension du système neuronal est pourtant fortement inspirée de l'approche cybernétique et donc ne se détache pas entièrement de l'approche béhavioriste : le système nerveux est un réseau dans lequel l'information est stockée dans les connexions synaptiques qui relient les neurones entre eux (et non au sein même des neurones). L'étude de ces interactions suffit à comprendre les mécanismes du système neuronal, nul besoin de maîtriser

les propriétés internes du neurone (que l'on peut voir comme la boîte noire), ni même de maîtriser le fonctionnement interne de chacune des connexions synaptiques ; trop nombreuses et complexes, elles semblent inaccessibles à l'investigation scientifique.

Les neurologues et psychologues McCulloch et Pitts ont développé en 1943 une modélisation simple et formelle du réseau de neurones à partir de cette explication neurologique de la cognition (McCulloch et Pitts, 1943 : 115-43). L'encodage des stimuli du monde réel est justifié par l'inspiration biologique de la modélisation qui utilise la métaphore du neurone afin de créer un programme adaptatif. Dans ce modèle, le neurone a une valeur d'entrée (soit 1, soit 0), auquel un poids est assigné qui déclenche l'activation des connexions synaptiques si la somme de ces poids dépasse un certain seuil. Dans ce cas, la valeur en sortie du neurone sera de 1, dans le cas contraire elle sera de 0. Le réseau de neurones artificiel corrige progressivement ses poids de connexions jusqu'à obtenir un poids qui permet de minimiser l'erreur produite en sortie (Mermillod, 2016 : 26). Autrement dit, selon ce modèle « bio-inspiré », les connexions entre neurones se créent ou se renforcent sous l'effet de l'activation simultanée de ces deux neurones ; effet induit par l'environnement. Ce sont les connexions synaptiques entre les neurones qui créent de la valeur et non l'entité du neurone en soi. La compréhension du réseau et donc du processus élémentaire neuronal permet, selon cette perspective, la modélisation du comportement de l'être humain. C'est donc selon l'approche comportementale du réel venant du père de la cybernétique que la première modélisation de l'esprit humain a été développée. En effet, selon cette perspective, c'est bien dans la compréhension du réseau et donc du processus élémentaire neuronal que se trouve l'explication du comportement humain lors de l'activité cognitive. L'établissement de ce simple processus biologique permet la modélisation mathématique et ainsi la reproduction informatique.

La première machine « connexionniste » dite adaptative, le Perceptron, est développée en 1957 par Frank Rosenblatt. Le modèle du réseau de neurones artificiels alors utilisé ne pouvait résoudre des fonctions non linéaires à la différence du système cognitif humain. Cette limite computationnelle mise en évidence par Minsky et Papert en 1969 marqua la fin de la recherche prenant le système neuronal sous l'approche béhavioriste comme assise à l'étude de l'esprit pour laisser place à l'analogie computationnelle de l'IA symbolique inspirée des théories cognitivistes (Minsky et Papert, 1972). C'est sur un profond désaccord ontologique que la recherche en IA prit un tout autre tournant ; la question technique (non-linéarité des fonctions) faisant seulement office de preuve aux failles de l'approche connexionniste.

### 2.1.2. L'IA symbolique

C'est à l'université prestigieuse de Dartmouth aux États-Unis que s'organise le séminaire qui marquera la naissance de l'IA à l'été 1956. L'atelier de recherche fut lourdement financé par la fondation Rockefeller (Bouée et Roche, 2017 : 23). C'est le mathématicien et informaticien John McCarthy qui fut à l'initiative de ce projet de recherche en collaboration avec Marvin Minsky (informaticien spécialiste des sciences cognitives), Nathaniel Rochester (co-concepteur du 701, le premier ordinateur d'IBM) et Claude Shannon (fondateur de la théorie de l'information). Une dizaine de chercheurs de disciplines diverses, mais de courants de pensée proche (la cybernétique, traitement de l'information, théorie des automates, modèle de prise de décisions, réseaux neuronaux formels) fut invités à participer à la création d'une nouvelle discipline qui rassemblerait ces différents domaines dans le but de se concentrer sur le développement d'une IA. Une vingtaine y participa, parmi eux des scientifiques de renom tels que Bigelow (collègue de Wiener), Mackay (spécialiste de la théorie de l'organisation du cerveau), Solomonoff (pionnier de l'apprentissage automatique), Newell et Simon (initiateurs du IPL, un langage de traitement de l'information, du Logic Theorist et GPS (General Problem Solver), les tous premiers programmes informatiques s'apparentant à l'IA.) C'est lors de ce séminaire qu'est utilisé pour la première fois le terme Intelligence Artificielle. McCarthy choisit ce nom pour clarifier sa pensée dans une idée de distinction et de neutralité. Il voulait se distinguer des études des automates donc il trouvait les théories trop étroites, et se dissocier de la cybernétique. L'ambition cybernétique d'une machine capable d'un ajustement adaptatif des entrées et des sorties lui semblait trop limitée (Benbouzid et Cardon, 2018 : 185). Le projet des pionniers de l'IA dite « symbolique » est de doter la machine d'une capacité de raisonnement. L'intérêt est porté sur le processus cognitif et l'acquisition de connaissances (objet de recherche propre au courant de pensée cognitiviste) plutôt que sur le comportement humain. En opposition au courant béhavioriste, le cognitivisme s'intéresse à l'étude des processus mentaux se déroulant dans la *boîte noire* entre l'environnement (stimuli pour le béhavioriste et entrée pour le cognitiviste) et le comportement (réponse pour le béhavioriste et sortie pour le cognitiviste). Selon cette approche, la cognition (qui constitue l'ensemble des états mentaux) n'est pas réductible au simple niveau neuronal (Benbouzid et Cardon, 2018 : 186). Le cognitivisme emprunte une approche computationnelle de l'esprit. C'est pourquoi ce courant est associé de près au computationnalisme (en opposition au connexionnisme) ; théorie qui utilise la métaphore de l'ordinateur pour affirmer que l'esprit humain est un système de traitement d'informations. Traitement, stockage, entrée (input), sortie (output), etc. : l'analogie avec

l'ordinateur est très claire aux vues de la terminologie appliquée aux processus mentaux dans la perspective cognitive.

Dans l'invitation formulée par McCarthy au séminaire de Dartmouth, « Proposal for the Dartmouth Summer Research Project on Artificial Intelligence » (McCarthy, Minsky, Rochester et Shannon, 2006), les chercheurs de l'IA symbolique (le premier véritable courant de la discipline autrement appelé IA classique) exposent leur aspiration :

« The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. » (McCarthy, Minsky, Rochester et Shannon, 2006 : 13)

À leurs yeux, il suffit de décrire les fonctionnalités de l'intelligence assez précisément pour ensuite les simuler à la machine et ainsi accéder à une « intelligence » artificielle aussi, sinon plus, performante que l'intelligence naturelle humaine. L'IA symbolique se considère comme science de l'esprit de la machine. Selon les chercheurs porteurs de ce courant, l'acte de penser revient à calculer des symboles qui ont une réalité matérielle et une valeur sémantique de représentation. À la différence de la métaphore neuronale connexionniste, il ne suffit pas d'identifier le processus élémentaire du raisonnement, mais d'établir les règles de représentations internes qui abritent les connaissances en tant que telles pour les insérer dans le programme informatique afin qu'il ait accès au monde réel. L'information est donc associée à des symboles qu'il convient de maîtriser afin de donner à la machine des capacités abstraites et logiques.

« On this basis we shall say that an entity is intelligent if it has an adequate model of the world (including the intellectual world of mathematics, understanding of its own goals and other mental processes), if it is clever enough to answer a wide variety of questions on the basis of this model, if it can get additional information from the external world when required, and can perform such tasks in the external world as its goals demand and its physical abilities permit. » (McCarthy et Hayes, 1969 : 466)

Pour qu'une entité (naturelle ou artificielle) soit intelligente selon McCarthy, il suffit qu'elle ait un modèle de représentation adéquat du monde ; d'où la nécessité pour la recherche en IA d'établir ce modèle de représentation afin de l'intégrer au programme. Nous verrons au chapitre II de façon plus approfondie sur quel postulat de l'intelligence humaine se base cette approche. Concrètement, l'idée est de placer des règles générales de connaissances permettant de manipuler des représentations de hauts niveaux à l'intérieur des ordinateurs à travers les programmes d'IA (Benbouzid et Cardon, 2018 : 186). La succession d'instructions agissantes est régie par des algorithmes complexes préalablement codés par les programmeurs. Non

seulement les règles de connaissances ou les informations stockées en mémoire (entrées) doivent être préalablement définies et calibrées par le programmeur, mais les résultats ou solutions (sorties) doivent aussi être anticipés par celui-ci pour que le système fonctionne, c'est-à-dire trouve celle appropriée. L'apprentissage en IA symbolique correspond à la capacité d'exécuter des actions sur la base de ce qu'on lui a inculqué. En d'autres termes, on explique au système de l'extérieur ce qu'il doit faire dans une situation donnée et le système reproduit ; c'est ainsi qu'il apprend. Tout comme nos parents et nos professeurs nous ont inculqué des valeurs, des façons de se comporter dans telles et telles situations que nous reproduisons. Du point de vue de l'IA symbolique, c'est ainsi que nous nous construisons. Dans cette perspective, l'esprit n'a pas un accès direct aux connaissances du monde réel. Ces représentations internes du monde doivent être décrites et organisées pour que l'esprit puisse s'en servir. Pour Herbert A. Simon, alors que les lois générales d'un système sont simples à intégrer, c'est l'environnement extérieur qui est complexe et pourtant non-négligeable à saisir, car le programme doit pouvoir adapter son comportement à celui-ci s'il veut réagir « intelligemment » (Simon, 2008 : 21).

Simon et Newell représentent la contribution la plus importante à l'efflorescence de l'IA et sont les fondateurs de ce qui est considéré comme les premiers programmes d'IA : le Logic Theorist (LT) et le General Problem Solver (GPS). Leurs intérêts se focalisent sur la résolution de problème et le raisonnement symbolique (Pomian-Saatchian, 1987 : 9). L'ambition de ces chercheurs est de réaliser une transcription des habiletés intellectuelles en entités quantifiables ; autrement dit, entreprendre une modélisation de l'activité cognitive en rassemblant l'ensemble des processus de logiques formelles. La recherche scientifique en IA s'opère par le biais d'une modélisation de l'intelligence. « Un modèle, en effet, est défini comme est un ensemble de règles permettant de reproduire un comportement observable. » (Robillard, 2019 : 15) Pour élaborer le modèle de l'intelligence, il s'agit de traduire la notion d'intelligence sous forme de langage formel. Contrairement au langage naturel, ce type de langage est formé d'un ensemble de symboles mathématiques obéissant à des règles strictes, ou « normes de *validité* » appartenant au modèle logico-mathématique (Robillard, 2019 : 15). Une même formule en langage formel peut être utilisée dans plusieurs modèles puisque la particularité de ce langage est qu'il ne tient pas compte de la sémantique ce qui facilite la manipulation et la transformation de l'énoncé de base. « (...) dans le domaine particulier de la logique mathématique, le modèle est un champ de réalisation de ce qui est exprimé par un système formel. », nous précise Emmanuelle Bulle dans l'article « Les modèles formels et l'explication en sciences sociales. » (Bulle, 2005 : 6) Ce modèle informatique offre une représentation symbolique d'un phénomène

empirique. Par la modélisation, l'intelligence est alors mobilisée en tant que langage de programmation formalisant les algorithmes. « Le modèle de l'intelligence dont il est question ici est construit à base d'algorithmes et la représente comme une structure computationnelle. » (Robillard, 2019 : 16) Ainsi chiffrée, l'intelligence est scientifiquement et informatiquement exploitable et peut donner lieu à la recherche en IA. C'est pourquoi la définition de l'intelligence en IA est essentielle à saisir, car les applications concrètes de ce domaine technoscientifique reposent sur une représentation sociale spécifique au sein d'un univers épistémologique particulier, c'est ce que nous verrons au chapitre II à travers l'exploration du concept de résolution de problème, objet de recherche principal des travaux de Newell et Simon.

Pour Minsky, nos « actions mentales » sont comme de petites machines fonctionnant ensemble l'une après l'autre selon un certain processus (Bouée et Roche, 2017 : 29). Il suffirait donc de reproduire artificiellement ce processus constituant l'activité cognitive qui représenterait l'intelligence humaine pour que la machine soit dotée à son tour d'intelligence. On retrouve la conception du modèle informationnel selon laquelle le système global dans ses relations et échanges d'informations prévaut au-delà des spécificités et fonctions particulières du cerveau. Seulement, si le fonctionnement de l'esprit est purement mécanique et que donc le cerveau est une machine au même titre que l'ordinateur, le degré quantitatif de la complexité de cette machine diffère grandement. L'intelligence serait la capacité de résoudre des problèmes de hautes complexités à partir d'informations symboliques traduites en langage binaire.

« Il n'y a pas la moindre raison de penser que les cerveaux sont autre chose que des machines comportant un nombre énorme de pièces fonctionnant dans le plus parfait respect des lois physiques. Pour autant que l'on puisse le dire, notre esprit n'est qu'un ensemble de processus complexes. Les grands problèmes proviennent de notre manque total d'expérience de machines d'une telle complexité, par suite duquel nous ne sommes pas encore prêts à réfléchir efficacement à ce qu'elles sont. » (Minsky, 1988 : 557)

Cette fascination des chercheurs envers le processus mental humain dans les années 60 empiète sur les avancées concrètes de l'IA. L'aspect pratique et la réflexion autour des applications possibles ont donc été délaissés en raison de ce dévouement à faire progresser les méthodes développées. Celles-ci servent davantage à justifier les postulats des chercheurs sur les secrets du raisonnement humain qu'à contribuer à la création d'une réelle IA. Les pionniers du domaine

étaient tout de même persuadés d'atteindre l'objectif d'une IA forte<sup>2</sup> rapidement sous la base des théories cognitivistes. Herbert Simon prédisait déjà en 1965 que les machines seraient capables d'ici une vingtaine d'années d'accomplir n'importe quelles tâches qu'un individu puisse exécuter (Simon, 1965 : 4). Les chercheurs de l'IA symbolique ont interprété cet échec par un manque de puissance de calcul des ordinateurs de l'époque ne permettant pas de délivrer des systèmes en concordance avec les hautes ambitions des chercheurs. En réalité, le problème provient davantage d'un écart trop important entre les règles de connaissances élaborées par les chercheurs et le monde réel. Les systèmes d'IA symbolique montrent des résultats satisfaisants face à certains problèmes spécifiques (ils sont tout à fait capables de jouer aux échecs par exemple), mais sont déplorables face à des tâches auxquelles l'individu est confronté quotidiennement (la conduite par exemple). Le programme n'est pas apte à faire face aux infinis et imprévisibles obstacles qui se placeraient sur sa route, tout simplement, car l'IA symbolique demande l'anticipation de toutes les possibilités intégrées préalablement dans le système. Ce problème est exposé par McCarthy dans l'article « Epistemological Problems of AI » (McCarthy, 1981), bien que sa conclusion reste en continuité totale avec la perspective des chercheurs de l'IA classique. Puisqu'une machine dotée de l'intelligence générale a besoin d'une vue générale du monde, les déficiences du programme sur leur propre vue du monde peuvent entraîner la faiblesse opérationnelle de la machine. Il ne remet donc pas en question l'approche de l'IA symbolique qui affirme qu'en attribuant à la machine les bases des connaissances représentant le monde par le biais d'une modélisation informatique, celle-ci serait alors dotée de raisonnement. Selon lui, pour que ce projet aboutisse, il nous faut une connaissance plus approfondie et adéquate de notre monde (McCarthy, 1981 : 16). Dans la perspective constructiviste qui est la nôtre, ce « problème » constitue plutôt un obstacle insurmontable.

Au vu des nombreuses critiques formulées par les philosophes démontrant en quoi la simplification du réel dans les modèles développés par l'IA symbolique ne peut ni fonctionner ni prétendre être proche d'une soi-disant « intelligence », de la faible puissance des calculs et

---

<sup>2</sup> L'IA faible représenterait le développement de programme conçu pour assister l'humain dans les tâches de la vie quotidienne. Ce serait plus un outil qu'une réelle duplication de l'esprit humain. L'IA forte représente, elle, représenterait un programme doté d'un esprit, d'une conscience (pour certains) et qui aurait la capacité de penser et d'agir aussi bien que l'être humain.(Nilsson, 2010 : 310)

des contraintes imposées par le département de la défense<sup>3</sup> ; les investissements industriels et gouvernementaux ont commencé à se faire rares. Coïncée entre les méthodes et les fantasmes et manquant de soutien financier à la recherche, l'IA va alors connaître une période d'hibernation à partir du milieu des années 70, désormais appelé « Hiver de l'IA ».

## **2.2. Concrétisation technologique.**

### 2.2.1. Renaissance de l'IA

#### *L'expertise intégrée à la machine*

Ce sont les « systèmes experts » qui marqueront la fin de l'engourdissement de l'IA à partir des années 80. Les fondements des programmes d'IA sont révisés de façon à répondre à la critique faite à l'encontre de l'IA classique. En effet, les connaissances ne sont plus conçues par les programmeurs eux-mêmes, mais par les experts d'un domaine spécifique pour chaque programme (Benbouzid et Cardon, 2018 : 192). L'idée est de simuler le travail d'un expert d'un domaine particulier : établir une proposition probabiliste à partir d'un échantillon de données précis d'une population. En s'ouvrant au monde des connaissances de spécialistes de toute autre discipline, l'IA s'ouvre sur l'extérieur lui permettant d'établir un lien réaliste avec le monde plutôt que d'inventer un univers abstrait. Le système expert est conçu dans l'idée de remplacer l'expert. Selon la définition de Corine Delmas dans l'ouvrage *Sociologie politique de l'expertise* : « L'expert se définit par rapport à la détention et à l'exploitation d'une compétence qui lui est reconnue et dont il tire, au moins partiellement, sa légitimité. » (Delmas, 2011 : 17) Faisant preuve d'autorité, le savoir expert constitue le référent du savoir véritable, seul savoir légitime. Par l'expertise intégrée à la machine, il devient compliqué de critiquer la représentativité des connaissances du monde réel des programmes d'IA ; cela reviendrait à critiquer la légitimité des experts. Le savoir de l'expert représente le véritable fondement du modèle technocratique de décision. En effet, toutes les décisions politiques, économiques et sociales sont remises à l'expertise des professionnels leur permettant une réelle implication en pratique, là où le profane ne pourrait intervenir. L'expertise scientifique faisait déjà preuve d'autorité depuis un moment, mais elle s'élève au rang de savoir dominant dans les années 1980

---

<sup>3</sup> Les chercheurs sont soumis aux contraintes strictes imposées par le Mansfield Amendment de 1969 qui stipule que toute recherche doit servir les besoins militaires.

(au point d'influencer les recherches de l'IA) entre autres en raison de l'attention croissante portée à la crise environnementale (Delmas, 2011 : 36). Il devient alors clair qu'une action publique doit être mise en place afin de faire face aux risques potentiels encourus par le dérèglement du climat. L'évaluation de ces risques dans un premier temps, puis la prise de décision politique pour les contrer ensuite est remis aux mains des experts de domaines variés. Seules capables de faire face aux défis de notre civilisation ; c'est à la science et à la technique que l'on remet cette responsabilité. L'ampleur du problème environnemental est si importante et insaisissable que l'exclusivité du pouvoir de l'expert n'est pas remise en question par le citoyen, dont la déresponsabilisation politique lui convient. Le développement et le succès des systèmes experts s'expliquent par ce contexte social qui valorise le savoir expert au-delà de tout autre savoir.

Les systèmes experts fonctionnent par propositions conditionnelles. Les connaissances sont transformées en propositions déclaratives formulées dans un langage le plus naturel possible utilisant l'inférence statistique. Avec l'apparition de ces nouveaux programmes, on passe d'un programme qui utilise un raisonnement logique (IA classique) à un programme utilisant un raisonnement heuristique. L'heuristique est une méthode de résolution de problèmes qui permet de fournir une solution réalisable rapidement en l'associant à une classe de problèmes déjà identifiés sans analyser le problème dans les détails.

« Naturellement, l'objectif final doit être de trouver une véritable conclusion. Mais, que les logiciens ou les puristes le veuillent ou non, le chemin de la vérité passe principalement par des approximations, des simplifications et des intuitions plausibles qui sont en fait fausses lorsqu'elles sont prises littéralement. » (Benbouzid et Cardon, 2018 : 193)

C'est donc par une approximation qu'on est capable de répondre à des questions par des exemples/ contre-exemples (négation d'une généralité) à partir d'une sélection d'hypothèses émises par le programme. Cette estimation ne peut plus être prise en charge par les règles des programmes d'IA symbolique. C'est à l'apprentissage automatique (champ d'étude de l'IA) que revient la charge de trier parmi ces hypothèses afin de sélectionner statistiquement la plus adéquate au problème.

L'accès à des calculateurs plus puissants et à des capacités de stockage plus importantes a permis aux systèmes experts de se faire une place notable dans l'avancée de la recherche en IA. Ces systèmes experts ont entraîné la spécialisation ; ce qui concrètement s'est retranscrit par le développement de sous-champs disciplinaires. Parmi ces sous-champs, on retrouve le

raisonnement et la représentation, le traitement du langage naturel, la vision par ordinateur et l'apprentissage automatique. Cette spécialisation de l'IA a aidé à contrer les critiques et à renforcer la légitimité des chercheurs. Elle a aussi permis de dissocier les recherches du nom « d'Intelligence Artificielle » ; terme entaché par les promesses non réalisées des premiers chercheurs. Cette stratégie a permis le retour des investissements. Les programmes sont tout de même d'une grande complexité et leur performance reste limitée. Les systèmes experts font cependant figure de pont (par le biais de l'apprentissage automatique) entre l'IA symbolique et le retour du connexionnisme.

### *L'apprentissage du côté de la machine*

L'apprentissage automatique se situe au carrefour de l'algorithme entièrement calibré par le programmeur dont les instructions sont rigides et les sorties anticipées, au système qui adapte son programme et ses résultats aux données qui lui sont transmises par un réseau d'entités élémentaires. Avec l'apprentissage automatique, on passe de la machine hypothético-déductive (déduire des faits empiriques particuliers à partir d'une hypothèse théorique générale) à la machine à induction (induire des lois générales d'observations de faits particuliers) tout en restant dans l'approche computationnelle (Benbouzid et Cardon, 2018 : 176).

Il est toujours question de reproduire ce qu'on pense savoir de l'être humain. Or, il devient clair pour les chercheurs que la question de l'apprentissage comme fondement de l'intelligence humaine doit être intégrée à la machine si l'on veut atteindre une IA digne de ce nom. Dans cette perspective, il est donc nécessaire que le programme, à l'image de l'Homme, se régisse lui-même puisque le processus d'acquisition de connaissances par expérience (soit l'apprentissage) fait preuve d'autonomie. L'apprentissage automatique introduit la notion d'*expérience* (acquisition des connaissances par la pratique selon la définition du psychologue Piaget), longtemps mise de côté par les chercheurs en IA, étant considérée comme l'élément propre à l'humain constituant ainsi le degré de distinction impénétrable par la machine. Rappelons que Wiener ne voit pas de différence de nature entre l'être humain et la machine. Le fait qu'il compare les deux entités montre bien qu'il les place toutes deux sur la même échelle. Néanmoins, une différence de degré persiste à ces yeux. Elle se trouve dans la capacité de l'humain à agir et prendre des décisions face à l'imprévu et à l'irrationnel (Wiener, Wall-Romana, Mopsik et Scholem, 2001 : 18). La machine, elle, n'en est pas capable. Il pense en revanche que ce n'est pas impossible que celle-ci atteigne un jour cette capacité, encore propre à l'Homme. Dans la même logique, la véritable IA (dite forte), sera atteinte selon Turing lorsque la machine sera capable d'appréhender l'environnement à l'instant  $t$  et de prendre des décisions

sans que l'homme ne la programme à cette fin précise (Turing, 1950 : 52). Or, c'est ce que l'apprentissage automatique affirme faire. En introduisant le concept d'apprentissage, d'expérience et d'adaptation à leurs algorithmes, c'est à cette distinction que s'attaquent les chercheurs en apprentissage automatique. Leurs algorithmes diffèrent donc des programmes classiques, car ils permettent au système d'adapter ses résultats sur la base des données traitées. Lorsqu'on parle d'apprentissage en IA, on entrevoit une métaphore de l'apprentissage chez l'humain. Seule une conception socialement construite et inscrite au sein du modèle informationnel de l'apprentissage peut être modélisée mathématiquement au sein d'un programme informatique. Les chercheurs en IA entendent ainsi par *apprentissage* la confrontation entre les résultats d'une action (ou ce qui est attendu d'elle) et l'amélioration de cette action en vue d'atteindre l'objectif attendu. L'action est évaluée par un nombre à l'atteinte de son objectif et plus le résultat est éloigné de l'objectif, plus on va renforcer les actions qui se rapprochent de cet objectif et réduire l'importance de l'action qui n'atteint pas cet objectif. Cette méthode fait écho à la théorie d'apprentissage par renforcement de Skinner : en récompensant un comportement, on tend à le faire se produire davantage, c'est ce qui s'appelle provoquer un stimulus de renforcement. L'apprentissage est donc dans ce sens une formule statistique d'optimisation. Selon cette perspective, l'être humain (ou le système) apprend de son expérience par un traitement de l'information composé d'éléments physiques concrets et calculables (les données). C'est le cumul quantitatif de ces éléments qui constitueraient l'environnement de l'humain. L'Homme s'adapte à son environnement et se base sur son expérience plutôt que par des règles strictes prescrites à l'avance (programme algorithmique classique) afin de guider son comportement (émettre un résultat). Selon cette perspective, l'expérience de l'individu se limite à son environnement extérieur et ne prend en aucun cas compte de l'environnement social et psychique de l'individu. Si ces aspects internes venaient à être pris en compte dans les théories de l'apprentissage, il serait alors impossible d'appliquer ces mêmes théories à la machine.

En cohérence avec le modèle de raisonnement par induction utilisé en apprentissage automatique, c'est par l'intégration d'une multitude de données qu'on pourra induire des lois générales concernant l'esprit humain. La connaissance se trouve dans l'accumulation massive de données appartenant et représentant tous les citoyens (Benbouzid et Cardon, 2018 : 177). La quantité d'information au-delà de sa qualité devient la réelle autorité. La frontière entre le savoir scientifique et le savoir profane s'estompe alors dans le vortex du Web. Le savoir expert perd alors en crédibilité et ne constitue plus le savoir dominant. La confiance aveugle autrefois accordée à l'expertise est aussi remise en question au vu du manque de résultats concrets pour

répondre aux grands conflits de notre temps. Nombreux reprochent à l'expertise scientifique de servir bien souvent des intérêts autres que ceux du bien commun (Delmas, 2011 : 87). L'utilisation systématique de l'argument de l'expertise pour tout débat et tout enjeu politique et économique et toutes décisions d'actions publiques sans résultats à la clé délégitime l'autorité du savoir expert.

### 2.2.2. L'apprentissage profond : retour du connexionnisme

Très peu représentatifs des chercheurs en IA, quelques-uns persistent à emprunter la voie connexionniste dans la conduite de leurs recherches. Parmi eux, Yoshua Bengio (chercheur à l'UdeM), Yann LeCun (directeur du laboratoire de recherche en IA de Facebook) sous la direction de Geoffrey Hinton (Google AI) s'intéressent dans les années 90 à un domaine alors abandonné : les réseaux de neurones artificiels (Benbouzid et Cardon, 2018 : 177). En reprenant les travaux de McCulloch et Pitt, combiné à la puissance de traitement plus importante que jamais et aux méthodes d'apprentissage automatique, ils sont capables de concevoir des algorithmes ayant des résultats spectaculaires donnant vie à des applications concrètes. Le développement de ces recherches donne lieu à l'apprentissage profond (en anglais, Deep Learning). Ce programme permet de réaliser des fonctions non linéaires (Mermillod, 2016 : 59). La grande limite computationnelle de l'approche connexionniste alors franchie, l'approche peut légitimement faire son retour. L'apprentissage profond est une méthode développée sur la base de l'apprentissage automatique. Le but est d'optimiser la prédiction à partir d'un échantillon très important de données traitées de façon brute, c'est-à-dire sans passer par une modélisation explicite comme on retrouve en méthode computationnelle (IA symbolique). En apprentissage profond, le système traite les données par un réseau de neurones artificiels multicouches. Dans ce programme chaque couche prend en entrée la sortie de la précédente (Mermillod, 2016 : 60). Par cet apprentissage à plusieurs niveaux d'abstraction, les entrées peuvent être intégrées en profondeur et ainsi être analysées de façons plus rigoureuses donnant des résultats aussi impressionnants qu'inattendus. Cette technique n'a pourtant pas connu un succès immédiat et fut lourdement contestée. Il était difficile de mettre de côté des décennies de recherches sophistiquées au profit d'une méthode élaborée aux prémices de l'IA sous preuve d'aucun nouveau fondement théorique véritable (nous y reviendrons dans le chapitre II). Cela dit, dépendant majoritairement de la quantité d'entrées à traiter (Bengio, Courville et Vincent, 2012 : 1), les programmes d'apprentissages profonds ont finalement percé à partir de 2010 grâce à l'extension du Web mettant à disposition des milliards de données par le biais de notre

utilisation quotidienne d'Internet. Les sorties de notre utilisation du Web (les cliques, les likes, les partages, les mots-clés, les photos, etc.) constituent les entrées des machines d'apprentissage profond (Benbouzid et Cardon, 2018 : 214). L'objectif global est d'implanter dans le système un ensemble de données censées représenter par leur multiplicité et leur valeur le monde environnant. Avec des résultats de plus en plus remarquables, il était impossible pour les investisseurs de ne pas y apporter une attention particulière. Les applications les plus utilisées au quotidien découlent de l'apprentissage profond : système de reconnaissance vocale comme Siri développé par Apple, la traduction automatique de Google et les voitures autonomes. Ces nombreuses applications d'optimisations ont suscité des investissements privés (notamment de la part des GAFAs, Google, Apple, Facebook, Amazon), et publics (gouvernements et universitaires) colossaux.

L'ambition fondatrice de l'IA repose sur la volonté de produire une créature à l'image de l'Homme. À la base d'un simple projet idéologique donc, l'IA doit sa concrétisation en tant que domaine de recherche scientifique à l'avènement du paradigme informationnel. Alors qu'à travers le paradigme cartésien l'intelligence se concevait à travers le jugement philosophique de ce qui est vrai et faux. Ce n'est qu'au sein du modèle informationnel introduit par la cybernétique que l'intelligence dans sa représentation informationnelle et opératoire a pu être pensée simuler à la machine. Le fonctionnement de la pensée humaine et de l'être humain se comprend au sein du monisme informationnel de la même manière que le fonctionnement de la machine. Autrement dit, les mêmes théories s'appliquent pour ces deux entités à partir du moment où l'on conçoit l'être humain et la machine dans la même catégorie ontologique. Bien que l'IA n'aurait pu être conçue en dehors du modèle informationnel, le développement de la discipline et de ses programmes informatiques s'est fondé sur des approches et méthodes théoriques qui dépassent celles formulées par la cybernétique. Des systèmes d'IA symbolique provenant de l'approche computationnalisme associée au cognitivisme, aux systèmes d'apprentissages profonds utilisant les méthodes de réseaux de neurones artificiels, en passant par une longue période d'hibernation, l'IA a réussi à se frayer une place aux côtés des disciplines scientifiques ayant pour objet de recherche l'étude de l'esprit tout en s'imposant sur le marché des nouvelles technologies.

Maintenant que nous avons exploré par un retour sociohistorique les fondements de l'IA de ses fondations idéologiques les plus profondes à sa concrétisation technologique ; il convient

désormais de se pencher sur la notion d'intelligence et d'éclairer sociologiquement le concept tel qu'il est compris en IA.

## Chapitre II.

### La notion d'intelligence.

L'IA (pas seulement en tant que domaine de recherche, mais en tant que projet idéologique) est révélatrice d'une conception particulière de l'être humain et de son intelligence. Concentrer notre attention sur cet objet de recherche précis (l'IA), permet alors de soulever une réalité sociale plus large. Nombreux sont ceux qui cherchent à se saisir des mécanismes généraux de l'IA, des applications qu'elle engendre et des enjeux qui en émanent. En revanche, peu conteste la notion de *l'intelligence*. Prise pour acquise, aucune définition rattachée à cette notion ne résiste pourtant à l'épreuve du temps et des sociétés qui la traverse. L'ambition de transmettre l'intelligence à la machine est toutefois la preuve que ce concept est central dans nos sociétés. La visée de ce chapitre est de soulever la définition de l'intelligence prise pour acquise par les chercheurs en IA. Nous tenterons de démontrer par l'entremise de ce chapitre en quoi l'IA s'est développée sur la base d'une représentation réductrice de l'humain et de son intelligence sans jamais que les fondements de cette assise ne soient évoqués tant ils sont dissimulés aux confins des représentations sociales. D'autant que ce « paupérisme épistémique » (Robillard, 2019 : 3) n'est pas sans effets : de réels choix scientifiques en découlent et entraînent la mise sur le marché d'applications concrètes qui guident et transforment notre quotidien soutenu par une communication faisant preuve de véritable économie de la promesse. C'est pourquoi une réelle réflexion de fond est de mise, sur quoi compte se pencher ce chapitre.

Nous tenterons de démontrer dans une première partie que l'intelligence est un concept construit. Nous expliquerons notre prise de position en faveur d'une approche constructiviste avant d'éclairer le concept à la lumière des représentations sociales. Il n'est pas question d'assigner une nouvelle définition au concept, mais de soulever celle qui légitime le projet de transférer l'intelligence à la machine. Étudier la définition de l'intelligence à travers l'IA non seulement permet d'établir un portrait des représentations de l'intelligence dans les sociétés occidentales modernes, mais elle permet aussi de repenser la définition de l'être humain. D'une perspective philosophique à une perspective scientifique, nous étudierons ensuite la progression de la construction du concept d'intelligence au fil des approches historiques majeures portées par les grands penseurs de notre civilisation de Aristote à Wiener, en passant par Descartes et Darwin ; sans quoi le projet de l'IA n'aurait pu être pensé. Il convient après ça d'éclairer le concept tel qu'il est compris en IA. C'est pourquoi nous consacrerons la deuxième partie à

l'analyse des transformations de la représentation de l'intelligence de l'avènement de la cybernétique à aujourd'hui. À chaque stade majeur de l'IA, une définition de l'intelligence différente est tenue pour acquise par les chercheurs qu'il est possible de soulever au regard des résultats sortant des programmes à chaque période. Nous rendrons compte ensuite de la conception réductrice de l'Homme et de son intelligence sur lequel l'IA s'appuie et par laquelle elle est rendue possible. Dans une troisième et dernière partie, nous étudierons notre objet de recherche à travers les phénomènes d'artificialisation de l'intelligence humaine et naturalisation de l'intelligence artificielle. Nous montrerons en quoi l'IA n'est envisageable que par la déconstruction du dualisme nature/ technique qui non seulement légitime le projet de rendre la machine intelligente, mais surtout justifie la possibilité d'une telle démarche. Nous démontrerons par la suite en quoi ce phénomène se retranscrit à travers l'idéologie transhumanisme, courant de pensée intellectuel et scientifique visant l'amélioration de l'Homme par les moyens des technosciences.

## **1. L'intelligence : un concept construit...**

Il paraît clair à la suite du chapitre I que l'IA émerge de sociétés qui considèrent l'intelligence comme la qualité déterminante de l'être humain. Utilisée quotidiennement par tout le monde, souvent pour juger de la valeur intrinsèque d'un individu, cette notion lourde de sens doit être sociologiquement éclairée à la lumière des représentations sociales afin de comprendre ce que l'on entend réellement par « Intelligence Artificielle ». Pour cela, il convient de se demander : quelle est la définition de l'intelligence ? En s'attardant sur cette question, on réalise très rapidement que la réponse dépend d'où, quand et à qui l'on pose la question. L'ambition de ce chapitre est d'analyser les définitions en évolution de l'intelligence tenues pour acquises par les chercheurs en IA au fil des courants qui la traversent. Tandis que différentes représentations du concept subsistent au sein du même cadre informationnel ; il existe aussi en dehors de ce cadre une multitude d'autres conceptions de l'intelligence. Le but de notre démarche n'est pas d'établir une définition sociologique de l'intelligence qui ne ferait qu'ajouter une de plus à la liste et n'éclairerait pas davantage le concept. L'objectif est plutôt de comprendre pourquoi il n'y a pas de définition univoque et consensuelle de l'intelligence. Il n'est donc pas question de réfuter les définitions proposées de l'antiquité à aujourd'hui, mais

de développer un postulat selon lequel toutes les conceptions de l'intelligence sont « vraies » dans leur cadre de représentation respectif.

### 1.1. ... à la lumière des représentations sociales

« Is there such a thing as intelligence? »

Mugny et Carugati

Gabriel Mugny et Felice Carugati à travers le livre *Social Representations of Intelligence* (Mugny et Carugati, 2009) démontre en quoi l'intelligence est une construction sociale. Selon eux, « l'intelligence » est un concept qui peut être compris seulement si les fondements psychosociaux des représentations qui lui sont associées ont été explorés. D'après leur analyse psychosociale, « l'intelligence » est une construction mentale de nature sociale. Ainsi, la signification de la notion varierait dépendamment de l'individu, du groupe social et société qui transmettent les représentations sociales (Mugny et Carugati, 2009 : 155). L'intelligence est comprise dans ce livre comme un phénomène social dont la construction historique et collective doit être analysée (Mugny et Carugati, 2009 : 163). Certes, centrale dans la société occidentale contemporaine, l'intelligence ne serait qu'un concept subjectif ; aucune définition univoque ne peut donc lui être assignée.

Cette perspective fait écho aux travaux de Berger et Luckmann dans la *Construction sociale de la réalité* (Berger et Luckmann, 1998). Ces deux sociologues de la connaissance nous invitent à appréhender la réalité comme une construction sociale. Selon l'approche constructiviste préconisée par ces derniers, la représentation que chacun se fait de la réalité est socialement construite à partir de sa socialisation. La conception de la réalité de l'individu (qu'il tient pour acquise) ne serait donc pas le reflet d'une réalité objective, mais le produit de l'intériorisation de normes et valeurs à laquelle l'individu est exposé en interaction avec son environnement social. Tout comme notre réalité nous semble être objective et universelle, la conception que l'on se fait de l'intelligence nous paraît être la seule qui vaille. De plus, elle constitue un élément significatif de l'identité d'un individu (Mugny et Carugati, 2009 : 159). En effet, les individus sont catégorisés en fonction de leur intelligence selon les critères établis et sélectionnés dans une représentation spécifique. Cette catégorisation n'est pas sans conséquence puisqu'elle gouverne nos jugements et comportements envers autrui (Mugny et Carugati, 2009 : 34). Les représentations sociales de l'intelligence opèrent sans que l'on ait toujours conscience qu'elles existent ; c'est pourquoi il est fondamental de les clarifier. Pour cela, il s'agira d'adopter une perspective constructiviste en accord avec le postulat de Mugny

et Carugati : toute définition de l'intelligence ou tentative de mesure dépend des représentations spécifiques à chaque société résultant de quels comportements, attitudes et processus mentaux sont valorisés (Mugny et Carugati, 2009 : p.23). Appliquée à cette approche, l'intelligence n'est autre qu'un concept construit. Ce sont les conditions de cette construction qui nous interpellent. Aucune des définitions rattachées à la notion d'intelligence n'est fixe dans le temps et dans l'espace, mais toutes dépendent des représentations sociales qui lui sont rattachées. Étudier un concept par le biais de cette approche consiste à analyser par quels processus les individus construisent symboliquement leur réalité qui recouvre leur conception des phénomènes. Ainsi, au regard des représentations sociales, l'intelligence artificielle ne constitue pas simplement un système doté d'intelligence. C'est un programme informatique doté de fonctions et mécanismes que les chercheurs en IA vont appeler « intelligence ». C'est leur représentation sociale spécifique de l'intelligence que les chercheurs en IA tentent de reproduire artificiellement.

Une étude dirigée par le professeur de psychologie et science de l'éducation Pierre R. Dasen en 1984 témoigne de l'importance de concevoir l'intelligence comme un concept construit sous l'angle des représentations sociales qui diffère d'une culture à l'autre (Dasen, 1984). Cette étude s'appuie sur le modèle interactionniste d'adaptation développé par le psychologue Piaget selon lequel l'intelligence est comprise comme une adaptation à l'environnement : « Un acte d'intelligence suppose [...] une régulation énergétique interne (intérêt, effort, facilité, etc.) et externe (solutions recherchées). » (Piaget, 1988 : 163) L'auteur cherche à étendre cette théorie piagétienne. Selon lui, si l'on part de cette conception de l'intelligence, il serait alors logique que les développements cognitifs des individus divergent selon les cultures. Chaque société cultive ses propres normes et valeurs qui déterminent les critères d'une personne dite « intelligente ». Dasen illustre cette thèse à travers la représentation de l'intelligence des Baoulés, un peuple de Côte d'Ivoire. Chez eux, quelqu'un d'intelligent — n'glouèlè en Baoulé (Dasen, 1984 : 426)— ce n'est pas un individu capable de résoudre un problème en se référant à son raisonnement logique et mathématique (ce qui est évalué dans les tests de QI et qui montre que c'est bien la sorte d'activité intellectuelle valorisée dans notre société). Un individu intelligent chez les Baoulés est un individu qui a le sens des responsabilités familiales et de la communauté et qui fait preuve de respect envers autrui : « he readiness to carry out tasks in the service of the family and the community » (Dasen, 1984 : 425). C'est une intelligence que l'auteur qualifie de « sociale » *a contrario* de l'intelligence « technologique » propre aux sociétés occidentales. Selon la définition de l'intelligence des Baoulés, reproduire l'intelligence humaine de façon artificielle est tout simplement inconcevable puisque les facteurs constitutifs de l'intelligence ne sont ni rigides ni

opérationnels. Cette étude démontre que l'intelligence est socialement construite et profondément inscrite dans un paradigme bien spécifique.

## **1.2. ...au fil des grandes approches historiques**

La définition établie de l'intelligence dans le domaine de l'IA repose sur une perspective précise de l'être humain relatif au cadre de représentation sociale dans lequel il est pensé. La conception logico-mathématique de l'intelligence telle qu'on la retrouve en IA dépend de son cheminement à travers chacune des grandes approches historiques qui ont refaçonné notre rapport au monde et redessiné les contours de l'espèce humaine. Afin d'envisager la reproduction de l'intelligence au sein d'un support matériel artificiel, il est essentiel de concevoir le concept d'intelligence dans une perspective scientifique. Nous verrons pourquoi et comment ce concept, traversé par les grands cadres de références fondateurs de notre civilisation occidentale, est d'abord appréhendé comme un concept sacré, puis biologique ; sans quoi l'IA en tant que domaine de recherche scientifique n'aurait pu voir le jour.

### **1.2.1. L'intelligence hylémorphique d'Aristote**

Comme on a pu le voir dans le premier chapitre, la corporalité de l'être est le point fondateur de la nature de l'Homme au temps de la Grèce Antique. Mais où se situe l'intelligence dans cette représentation de l'être humain dont la beauté du corps est prédominante ? Pour Aristote, l'intelligence serait une puissance inorganique qui émanerait de l'âme (Demange, 2003 : 65). Selon la conception hylémorphique développée par Aristote qui considère que toute entité serait composée d'une matière (le corps) et d'une fonction (l'âme) associées, imbriquées et indissociables l'une de l'autre, le projet de reproduire artificiellement l'intelligence humaine n'est donc tout simplement pas concevable à l'antiquité. Même si les moyens techniques avaient été présents, la perspective d'unité du corps et de l'âme ne permet pas de concevoir l'intelligence séparée du corps humain prenant vie dans un autre support.

### **1.2.2. L'intelligence sacrée de Descartes**

L'évolution du concept d'intelligence prend un tout autre tournant avec l'introduction du dualisme cartésien. Descartes reconnaît l'existence des deux substances fondatrices de l'être humain : le corps et l'esprit. Le bouleversement épistémologique tient du fait que ces deux

instances sont, d'après lui, séparées et indépendantes (Gardner, Kornhaber et Wake, 1996 : 33). Le corps en tant que matière se situe du côté de la machine, tandis que l'esprit se trouve du côté du sacré. L'intelligence, présent au niveau de l'esprit, n'est donc dès lors plus associé au corps. L'esprit serait ce qui sépare l'Homme de l'animal et le distingue de la machine. Sacrée, l'intelligence n'est donc pas une fonction biologique. Elle n'est pas une instance naturelle au sens qu'elle n'est pas organique. Cette instance immatérielle est spirituelle et c'est pourquoi on ne peut l'expliquer que par le biais de la philosophie. Or, nous verrons qu'une perspective scientifique de l'intelligence est absolument nécessaire à l'établissement du projet de l'IA. On raconte qu'au XVII lors d'une promenade dans les jardins royaux français, Descartes se trouva face à des machines automatiques à allures humaines. Ayant la capacité de bouger et d'émettre des sons à l'image de l'homme, peut-on alors dire qu'elles sont douées d'intelligence ? Une question à laquelle Descartes aurait répondu négativement. Selon lui, aucune machine ne sera jamais capable de raisonner puisque qu'elles ne sont que matière, et que le raisonnement provient de l'esprit, de l'âme complètement séparée de la dimension matérielle (Gardner, Kornhaber et Wake, 1996 : 163). Si la représentation de l'intelligence inscrite dans le dualisme cartésien n'est pas suffisante à la consécration même de l'idée de l'IA telle qu'on la connaît aujourd'hui, elle a toutefois permis de légitimer l'idée selon laquelle cette entité autonome puisse se former indépendamment du corps humain ; ce dernier ne servant que de simple support. Ce dualisme corps et esprit est un des principes fondateurs des sociétés occidentales modernes sans quoi l'IA n'aurait pu être pensée.

### 1.2.3. L'intelligence biologique de Darwin

Afin que le projet de l'IA ne soit ne serait-ce que pensable, le rapport des Hommes au sacré a dû être démantelé. La théorie de l'évolution, en plus de bouleverser profondément et à jamais notre rapport au monde, marque un véritable tournant pour l'étude de l'intelligence. Formulée par Darwin entre 1859 et 1880, cette théorie exposée dans l'ouvrage *L'origine des espèces* démontre que toutes les espèces vivantes ont évolué au cours du temps à partir d'ancêtres communs ; une évolution due au phénomène de ce qu'il appelle la *sélection naturelle* selon laquelle les organismes les mieux adaptés à leur environnement sont ceux qui survivent et qui transmettent leurs gènes, entraînant de nouvelles mutations et ainsi l'*évolution* des espèces (Gardner, Kornhaber et Wake, 1996 : 39). Alors que l'explication de l'origine de l'Homme était d'ordre religieux depuis des millénaires, cette théorie émet pour la première fois une interprétation scientifique de la genèse du monde. Prédisposé à s'affranchir peu à peu des

fondements religieux grâce au développement de la science et de l'influence des Lumières qui remettent en question le dogmatisme religieux, le contexte est favorable à la réception de cette théorie controversée. La séparation de l'être humain et des autres animaux de par ses fonctions cognitives supérieures ne peut plus s'expliquer désormais par des présupposés théologiques. De ce fait, l'esprit ne peut plus être compris comme une instance appartenant au domaine du sacré. La séparation entre le corps et l'esprit est toujours présente, seulement le fonctionnement de ces deux instances peut désormais être étudié sous l'angle scientifique. L'héritage des déterminants constituant l'être humain ne se limite pas qu'aux éléments du corps, mais aussi à ses habiletés mentales. D'ordre génétique, l'intelligence est alors considérée comme organique. Le cerveau est l'organe porteur de l'esprit, de l'intelligence humaine. La théorie de l'évolution légitime l'approche de l'esprit dans une perspective non plus philosophique, mais scientifique.

### *La mesure de l'intelligence : le QI*

L'étude de l'intelligence dans une perspective scientifique a d'abord été abordée sous l'angle de l'observation de la distinction des intelligences. C'est par l'évaluation quantitative hiérarchisée que le critère de différenciation est mesuré. Assurément, afin de calculer l'intelligence d'un individu il faut initialement élaborer une définition de l'intelligence sur laquelle l'échelle d'évaluation peut s'appuyer. Or, la définition de l'intelligence a longtemps été liée à son mode d'évaluation (Andrieu, Anderson, Le Parvis, 1992 : 171). Le psychologue Edwin Boring disait « Intelligence is what the test test ». (Boring, 1961 : 211) Ce qu'on évalue à travers ces systèmes de mesures est démonstratif d'une représentation particulière de l'intelligence humaine et réciproquement alimente cette représentation. Les tests d'intelligence ont été élaborés dans l'optique d'établir une hiérarchie purement quantitative de l'intelligence. En effet, depuis les cent dernières années les méthodes d'examen de l'intelligence des psychométristes s'appuient sur des analyses statistiques (Gardner, Kornhaber et Wake, 1996 : 5). La première échelle d'intelligence est publiée par Binet et Simon en 1905. C'est à la demande du ministère de l'Instruction publique à Paris qu'ils élaborent cette échelle dans le but d'évaluer l'étudiant afin d'assurer une éducation adéquate aux enfants ayant des déficits mentaux (Binet et Simon, 2004). Cette échelle est démonstrative de la représentation de l'intelligence des deux psychologues et se fonde naturellement sur les normes de l'époque. C'est l'habileté de s'adapter aux circonstances, d'effectuer un jugement, de détenir des notions de sens commun pratique et la capacité d'initiative qui sont donc étudiés. « Nous avons voulu montrer qu'il est possible de constater d'une manière précise, vraiment scientifique, le niveau mental d'une intelligence, de comparer ce niveau au niveau normal, et d'en conclure par

conséquent de combien d'années un enfant est arriéré. » (Binet et Simon, 2004 : 336) Voilà la volonté des auteurs : pallier les failles d'un système subjectif d'évaluation du niveau intellectuel de la population faite par les psychologues et médecins selon des critères différenciés et leur perception personnelle de la situation. Tandis que l'évaluation standardisée est conçue pour déterminer *l'âge mental* de l'enfant, c'est le QI (quotient intellectuel) —comme son nom l'indique, le résultat d'une division— qui crée la véritable mesure. Introduit en 1912 par le psychologue Stern, le calcul de l'intelligence permet une standardisation d'autant plus généralisée, mais représente surtout le premier pas vers une modélisation de l'intelligence. L'âge mental divisé par l'âge réel et multiplié par 100 nous donne ce qu'on appelle le QI classique. Cette formule mathématique n'est pas concrète ; un âge est divisé puis multiplié par un nombre qui permet simplement une vision simplifiée par le pouvoir du chiffre. Le QI est censé exposer par ce chiffre les capacités intellectuelles d'un individu. N'ayant aucune signification algébrique réelle, elle légitime toutefois l'échelle de mesure par le moyen du calcul.<sup>4</sup>

Le QI classique laisse place en 1939 au QI standard. Le psychologue Wechsler propose une nouvelle échelle de mesure de l'intelligence. Le QI ne repose alors plus sur une simple division mathématique, mais sur une formule statistique. C'est le premier à introduire cette notion d'étalonnage encore utilisé aujourd'hui pour le calcul du QI (Gardner, Kornhaber et Wake, 1996 : 79). Par ailleurs, il est nécessaire de regarder sur quoi repose l'échelle de mesure. Rappelons que le QI repose sur un test d'évaluation des compétences intellectuelles d'un individu. Le test de Spearman introduit la notion du *facteur g* (le facteur général) qui rassemble un ensemble de variables censé représenter l'intelligence d'un individu. C'est ce facteur qui détermine enfin le QI général d'un individu. Le psychologue Thurstone émet une lourde critique envers cette manière d'appréhender la totalité de l'intelligence humaine. Le *facteur g* est pour lui non existant puisqu'il consiste à assembler des habiletés mentales indépendantes les unes des autres. Il n'y a pas de raisons empiriques assez viables qui justifient de quantifier des habiletés mentales dans le simple but de pouvoir émettre une formule mathématique peu rigoureuse. Il remet ainsi en question l'existence même d'une intelligence générale comme

---

<sup>4</sup> Dominique Cardon nous parle à travers *A quoi rêvent les algorithmes* de l'ébranlement du chiffre et des statistiques. Parmi de multiples autres indicateurs statistiques, le QI représente l'avènement de ce qu'il appelle *la société du calcul*. « L'objectif de ces indicateurs est moins de connaître le réel que de « conduire les conduites ». (Cardon, 2015 : 10) Utilisé initialement comme technique de gouvernement, la mesure statistique bouleverse notre rapport au monde et de soi. Aujourd'hui, on confie au calcul, aux statistiques, au chiffre le soin de nous diriger, de nous classer et d'ériger ainsi nos vies. C'est la preuve de notre pleine confiance envers les mathématiques. Le chiffre, au-delà de la pensée, représente le nouveau prisme de la rationalité.

entité mesurable. Dans une tentative d'explication multifactorielle, Thurstone élabore une nouvelle échelle dont les fondements sont toujours utilisés aujourd'hui. Le type d'épreuve composant cette évaluation nous indique précisément ce qui, aux yeux des spécialistes, compose l'intelligence humaine. Chacune des sections du test de mesure de l'intelligence est censée représenter un facteur d'habileté intellectuelle : le facteur spatial, le facteur de la perception, le facteur verbal, le facteur lexical, le facteur de la mémoire, le facteur numérique et le facteur du raisonnement (Thurstone, 1935). Ces facteurs composant l'échelle de mesure du QI constituent aussi les secteurs de recherche de l'IA. Autrement dit, il est de l'ordre de l'acquis en IA que ces sept facteurs soient les instances qui composent l'intelligence humaine. Or, ces « facteurs » ou habiletés mentales, supposés représenter l'essentiel de l'intelligence sont exclusivement des priorités qui tiennent des capacités logiques et mathématiques. Cette représentation particulière de l'intelligence qui est montrée à travers l'évaluation du QI met l'accent sur des aspects de la cognition aisément transposables en langage binaire et en formules mathématiques. Cette représentation logico-mathématique est traduisible, modélisable et donc possible de reproduire artificiellement.

Si le développement du système de mesure repose sur une représentation logico-mathématique de l'intelligence humaine ; réciproquement, les tests d'intelligence alimentent cette représentation. Il s'agit d'établir une standardisation par le biais de cette échelle métrique. Déjà, Binet et Simon mettaient un point d'honneur à regarder exclusivement l'état actuel de l'individu sans prendre en compte son environnement social et son expérience subjective passée (Binet et Simon, 2004 : 191). L'échelle de l'intelligence introduite par le calcul du QI participe ainsi à la représentation réductrice de l'être humain et de son intelligence. Les facteurs sélectionnés dans l'évaluation et qui constituent aussi les branches de recherche de l'IA correspondent aux aspects de l'être humain valorisés dans les sociétés occidentales contemporaines. Le raisonnement logico-mathématique semble représenter la valeur centrale à cultiver dans une société productiviste axée sur la performance et l'efficacité.

### *Les sciences cognitives*

L'étude de l'intelligence sous une perspective scientifique détache peu à peu sa définition de son mode d'évaluation. Dès lors, ce n'est plus la mesure de l'intelligence qui intéresse les chercheurs, mais son fonctionnement. L'étude de la pensée humaine est à ce terme le cheval de bataille des biologistes, psychologues, neurologues, linguistes et chercheurs en IA sous le nom d'une nouvelle discipline scientifique : les sciences cognitives. Caractérisée par son interdisciplinarité, la discipline prend vie lors d'un symposium sur le thème de la théorie

de l'information en 1956. Les fondements des recherches dans le domaine de l'IA se posent donc sur des bases théoriques de divers domaines, ce qui, en surface, légitime leur conception : une vision plus globale de l'être humain permettrait une vision plus juste de celui-ci. C'est en tout cas la thèse que formule Joëlle Proust, dans l'article « Redéfinir l'humain, pour une convergence des sciences de l'Homme. » (Proust, 2014) Seulement, ces domaines se rejoignent sur un point : leurs assises prennent forme à partir d'un même cadre épistémologique ; le modèle informationnel. Alors que les chercheurs pensent préconiser une vision globale, on peut se demander si ce n'est pas plutôt une vision unique qui ressort de cette « convergence des sciences de l'Homme ». (Proust, 2014 : 69). La conception de l'humain n'en est-elle pas ultimement réduite ? C'est ce que nous tenterons d'entrevoir dans la seconde partie de ce chapitre.

#### 1.2.4. L'intelligence informationnelle de Wiener

Comme nous l'avons vu au chapitre I, la cybernétique constitue évidemment la dernière rupture qui a ouvert la porte à une représentation de l'intelligence capable d'être retranscrite au sein d'une machine. L'esprit et le corps sont toujours deux instances séparées. Mais l'esprit n'est plus une instance sacrée, elle est de l'ordre du biologique. Le développement des sciences cognitives qui s'attèlent à découvrir les secrets du raisonnement humain de manière scientifique a été considérablement influencé par la cybernétique qui laisse place au modèle informationnel. Au sein de ce nouvel ordre de représentation, l'esprit se retrouve désormais du côté de la machine. L'intelligence est comprise comme un système informationnel du même ordre que tout autre système, être ou objet de ce monde. L'intelligence est une instance à part, certes d'origine biologique, mais qui peut être comprise dans son ensemble indépendamment de l'être humain en soi et de son corps, considéré comme simple support illusoire, c'est-à-dire secondaire puisque matériel. La vision informationnelle décristallise la place de l'esprit permettant le déplacement de l'information d'un support matériel à l'autre. Dans cette configuration, il est alors envisageable de transmettre cette instance informationnelle de façon artificielle dans tout autre support comme celui de la machine.

De plus, nous l'avons vue, le modèle informationnel opère une indifférenciation ontologique entre l'humain et la machine (Lafontaine, 2004, p. 43). Au sein du modèle unique informationnel, tout s'explique via un processus global pour toute entité humaine comme non humaine. Dès lors que les priorités internes sont négligeables, mais que seul le mécanisme absolu est pris en considération, on peut alors se permettre de reconnaître l'œuvre d'une même sorte d'abstraction en l'intelligence humaine et en l'intelligence artificielle. Selon cette logique,

les distinctions de nature entre l'esprit humain et le système informatique ne sont pas prises en considération ; seul le fonctionnement mécanique global et commun aux deux entités est étudié. C'est précisément cette indifférenciation analytique de l'entité naturelle et artificielle qui justifie le projet de l'IA. « Cette opération [le désir de construire un être à l'image de l'Homme] est rendue possible par une double réduction, de l'humain à l'intelligence, et de l'intelligence au traitement de l'information. » (Breton, 1995 : 138) À partir de l'avènement de la cybernétique introduite par Wiener, notre représentation globale du monde s'appuie sur le modèle informationnel. Selon ce précepte, la vie en elle-même est subordonnée au principe de l'information. La valeur de l'être humain résiderait donc en sa capacité à traiter l'information.

## **2. La représentation de l'intelligence en IA**

### **2.1. Une définition en évolution**

Il est clair que l'IA n'aurait pas pu exister dans un autre cadre épistémologique que celui du modèle informationnel instauré par la cybernétique. Cependant, l'IA s'est construite dès ses prémisses selon des conceptions de l'esprit humain en constante évolution provoquant des divergences quant à l'approche de recherche à adopter. Autrement dit, si les définitions de l'intelligence de Wiener à Yoshua Bengio en passant par Newell et Simon appartiennent toutes au même univers de représentation, celles-ci, traversées par des approches théoriques divergentes en évolution, divergent. La méthode, le programme, les entrées apprivoisées ne seront pas les mêmes dépendamment de la sortie du programme, soit de l'objectif de départ recherché. On cherche à travers l'IA à reproduire artificiellement ce qu'on pense caractériser l'intelligence. Le but de chacun de ces programmes est donc de générer un comportement « intelligent ». Si la rétroaction est la sortie escomptée de la machine cybernétique, cela signifie que l'on conçoit l'intelligence comme la capacité de rétroaction d'un système. Il en est ainsi à tous les stades de l'IA. La définition de l'intelligence se trouve dans la finalité recherchée de son programme informatique. Ainsi, pour les chercheurs en IA symbolique l'intelligence se résume en la capacité à résoudre des problèmes, et finalement, pour les chercheurs de l'apprentissage profond, l'intelligence se définit par l'optimisation en vue d'un objectif. Il convient alors d'étudier les transformations de la représentation de l'intelligence informationnelle de l'avènement de la cybernétique à aujourd'hui.

#### **2.1.1. La rétroaction**

Comme nous l'avons vu au chapitre 1, l'approche connexionniste appréhende le système nerveux comme un réseau de connexions dans lequel transite l'information. Concevoir l'esprit et donc le programme de l'IA sur la base de la métaphore neuronale fait écho à la pensée cybernétique. Rappelons que la rétroaction constitue le noyau de la pensée cybernétique. Or, la cybernétique est développée dans une perspective qui dépasse la simple ambition d'un développement scientifique. L'idée est de trouver une solution face aux dérives de ceux qui ont le pouvoir sur la société et d'ainsi, lutter contre l'entropie. Le traumatisme de l'holocauste pèse lourdement sur sa théorie. Selon Wiener, c'est l'échange informationnel qui a le pouvoir de s'y attaquer par le biais de la rétroaction. Dès lors, les premières machines connexionnistes sont conçues dans l'optique d'émettre un effet de rétroaction en guise de sortie. Le comportement d'un système ne serait que le résultat des interactions produites par des opérations élémentaires. Autrement dit, les finalités du système (naturel et artificiel) ne sont pas internes, mais produites par l'adaptation de son comportement au vu des erreurs qu'elle commet. L'intelligence est donc, dans cette perspective, la capacité d'adapter son comportement aux signaux de l'environnement constituant l'information ; la définition même de la rétroaction. Autrement dit, ce serait la capacité « d'orienter et de réguler ses actions d'après les buts visés et les informations reçues », selon la définition de la rétroaction de Céline Lafontaine (Lafontaine, 2004 : 46). Le processus de rétroaction constitue donc la définition cybernétique de l'intelligence. À la différence de la perspective behavioriste qui conçoit le comportement comme un simple réflexe conditionné, la notion de rétroaction intègre au comportement « intelligent » la possibilité d'ajuster la relation stimulus/ réponse en fonction des informations assimilées et des objectifs convoités. Cette différence constitue aussi la justification de la désaffectation de la loi de Hebb à l'approche behavioriste ; pas suffisant cela dit pour justifier son appartenance à l'approche cognitiviste. L'intelligence, dans cette conception, ne prend pas en compte la notion d'autonomie. Le processus d'intelligence permet seulement au système (naturel comme artificiel) d'entreprendre une action en réaction aux signaux de l'environnement auquel il s'adapte.

### 2.1.2. La résolution de problème

Nous le répétons, la représentation de l'intelligence sur laquelle repose le projet de l'IA ne peut exister en dehors du paradigme informationnel. L'IA symbolique assume cependant un net détachement avec la cybernétique par une divergence d'opinions quant à ces méthodes d'approches théoriques envers l'étude de l'intelligence.

L'évolution de la représentation de l'intelligence s'explique par la transition entre la conception behavioriste et cognitiviste. Les pionniers de l'IA symbolique dénoncent le réductionnisme du behaviorisme incapable de résoudre les grands questionnements concernant la complexité de la pensée humaine comme la mémoire, le langage, la capacité de résoudre des problèmes complexes, etc. (Nicolas et Ferrand, 2008 : 139). Selon eux, on ne peut étudier l'organisation de l'esprit en faisant l'économie de ce qui n'est pas observable à l'œil nu ; c'est pourquoi il est nécessaire de dévoiler les mécanismes internes présents. Rappelons que les cognitivistes conçoivent la cognition comme une manipulation de symboles effectuée sous un ensemble de règles ; les symboles constituant des aspects réels du monde. Il convient pour les chercheurs en IA symbolique de décrypter ces règles et de les modéliser afin de pouvoir les insérer dans l'ordinateur sous forme de programme algorithmique.

Le détachement de la représentation de l'intelligence avec le concept de rétroaction propre à la cybernétique est porteur d'une espérance humaniste ; en raison d'une volonté de sortie d'un réductionnisme assigné à l'étude exclusive du comportement, mais aussi conformément à l'ambition d'affirmer une distinction nette entre l'humain et l'animal (Nicolas et Ferrand, 2008 : 138). Autant l'indifférenciation ontologique entre la machine et l'humain ne semble pas poser de problèmes, autant l'assimilation de toutes espèces vivantes semble contrarier. Wiener concevait toute entité (animal, machine et humain) comme une organisation régit par le même processus informationnel. Les chercheurs de l'IA ne réfutent pas ce point théorique, mais souhaitent souligner la place de l'être humain au sommet de la pyramide hiérarchique des vivants. Ils insistent sur la différence de degré de complexité entre l'être humain et l'animal replaçant l'Homme en haut de la hiérarchie. Dans cette perspective anthropocentrique, l'intelligence est l'exclusivité de l'humain ; il convient donc de mettre en avant la complexité des processus cognitifs de l'être humain qui fait de lui une espèce à part. La simulation des processus cognitifs au sein de la machine doit donc être fondée sur ce qui fait la caractéristique de l'Homme et ce qui l'élève parmi les autres ; c'est-à-dire sa capacité de manipuler des représentations symboliques de hauts niveaux lui permettant de résoudre des problèmes d'une grande complexité. Le courant cognitiviste, et donc les chercheurs en IA

symbolique, conçoit ainsi l'intelligence comme la capacité de résoudre des problèmes complexes par la manipulation de symboles.

Il convient à présent d'analyser plus en profondeur ce que les chercheurs entendent par « résolution de problème » puisque l'éclaircissement du concept permet la compréhension de la représentation de l'intelligence en IA symbolique.

Les cognitivistes de l'IA identifient la résolution de problème comme étant le processus d'identification puis de mise en œuvre d'une solution à un problème en utilisant des méthodes génériques. Par problème, ils entendent toutes activités ou opérations « complexes » que nous entreprenons : empiler des cubes, assembler un puzzle, jouer aux échecs, élucider une énigme ou même démontrer un théorème. Le terme *complexe* est à saisir dans le sens de ce qui ne fait pas partie de l'acquis, du réflexe, du sens commun comme d'aller aux toilettes lorsque l'envie d'uriner nous prend. Par résolution, ils entendent la série de décisions en vue d'une solution menant à cette action. L'acte de résoudre un problème demande la mise en place consciente d'une stratégie. (Minsky, 1988 : 122) Le simple fait de se poser la question « comment vais-je atteindre mon objectif ? » suppose donc une situation de résolution de problème. Alors que l'être humain doit découvrir comment voler, l'oiseau, lui, ne se demande pas comment ni pourquoi il vole. Selon Minsky et ses compères cognitivistes, se serait l'aptitude d'identification de l'information à partir de rien et la mise en place consciente d'une stratégie visant un objectif (soit la résolution d'un problème) qui accorderait à l'homme sa distinction avec les autres espèces vivantes et qui lui donnerait par là son caractère singulier et dominant. « Ces animaux ne « résolvent » pas eux-mêmes ces problèmes : ils se bornent à exploiter les procédures présentes dans leur cerveau aux structures génétiques complexes »(Minsky, 1988 : 123); tandis que les êtres humains par le discernement de ces procédures ont la faculté de sélectionner celles qui seront les plus efficaces en fonction du problème encouru. Ils sont donc capables de s'adapter en exploitant leur raisonnement. Le projet idéologique de transmettre l'intelligence à la machine est très présent dans l'esprit des chercheurs en IA symbolique. La résolution de problème étant la cible principale de leur recherche, il paraît clair que ce concept caractérise à leurs yeux la définition même de l'intelligence. Elle serait donc le mécanisme mis en place par l'individu afin d'atteindre un résultat. Selon cette perspective, l'essence de l'intelligence se trouve dans le raisonnement logique. L'objectif est de donner à la machine des capacités abstraites et logiques semblables de celles mobilisées par l'Homme. La logicisation de la pensée permet de plus la mécanisation. L'intelligence réduite à un processus extrêmement mécanique est alors tout à fait intégrable à un système artificiel. C'est en faisant appel à la

logique mathématique que l'analyse des causes et conséquences permet de résoudre un problème. C'est ce que le psychologue Gardner appelle « l'intelligence logico-mathématique » et qui constitue, selon lui, qu'une partie seulement, de l'intelligence humaine. (Gardner, 2009 : 41) L'IA symbolique cherche ainsi à créer une machine hypothético-déductive, méthode scientifique consistant à formuler une hypothèse dont la validité est déterminée par la déduction de conséquences observables. « Le moyen le plus efficace de résoudre un problème consiste à savoir à l'avance comment le résoudre. » (Minsky, 1988 : 129) L'IA symbolique consiste donc, à identifier les mécanismes cognitifs présents dans la résolution de problème afin de les transférer à la machine pour qu'elle puisse, à la manière dont procède l'être humain, résoudre à son tour des problèmes semblables. On cherche à confirmer ou réfuter par les résultats de la résolution de problème un postulat représenté par les règles de représentations symboliques insérées dans le programme.

Il s'agit à présent d'étudier de plus près l'approche méthodologique employée par les chercheurs afin de transférer les procédures mentales mobilisées dans la résolution de problème dans la machine.

Newell et Simon développent deux programmes de résolution de problème. Le Logic Theorist, apparu en 1956, démontre la théorie de la logique formelle (Pomian-Saadjian, 1987 : 13). Le GPS conçu de 1957 à 1967 est, comme son nom l'indique plus général. C'est un programme de résolution de problème séparant les stratégies de résolutions de problèmes du type de tâches à exécuter (Pomian-Saadjian, 1987 : 15). Le but de ces deux programmes est de construire une machine qui résout des problèmes non algorithmiques, soit, capable de résoudre toutes sortes de problèmes. L'objectif final de la création de ces programmes n'est pas seulement l'acquisition de l'IA, mais plutôt le développement et la validation des théories cognitivistes psychologiques concernant l'intelligence humaine. C'est d'ailleurs ce qu'ils exposent en première ligne du livre considéré comme le manuel d'initiation à la recherche empirique de l'IA *Human Problem Solving* : « The aim of this book is to advance our understanding of how humans think. It seeks to do so by putting forth a theory of human problem solving, along with a body of empirical evidence that permits assessments of the theory. » (Newell et Simon, 1972 : 1) En effet, à partir de ces systèmes, Newell et Simon avaient dans l'idée de construire une théorie psychologique du traitement symbolique de l'information. Ils défendent la théorie selon laquelle le programme computationnel et l'être humain sont tous deux des systèmes de traitement de l'information puisque leur fonction principale est de résoudre des problèmes symboliques. « Humans, when engaged in problem solving in the kinds

of tasks we have considered, are representable as information processing systems. » (Newell et Simon, 1972 : 788) Si la théorie de la logique développée fonctionne à l'intérieur de la machine, alors la même théorie devrait s'appliquer pour l'Homme. L'ordinateur n'est plus qu'une métaphore, mais un outil permettant une comparaison détaillée entre les conceptions théoriques et les résultats empiriques à partir de la simulation. Par un raisonnement analogique, ils tentent de démontrer que si l'ordinateur est capable de simuler la pensée logico-mathématique de l'être humain, alors on peut affirmer que la pensée humaine fonctionne comme un ordinateur. La valeur démonstrative de cette démarche nous semble largement contestable. En admettant que le concept d'intelligence soit fixe et univoque, la reproduction du mécanisme global de celui-ci ne suffit pas à la reproduction de l'instance même de l'intelligence. D'où le questionnement très pertinent du sociologue Robillard : « si le modèle reproduit le mécanisme observé, en est-il pour autant une représentation adéquate ? » (Robillard, 2019 : 16). C'est en tout cas le postulat que proposent les deux chercheurs et sur lequel repose leur explication du raisonnement humain.

Avec les systèmes experts, le fondement de la définition de l'intelligence réside toujours dans la capacité de résolution de problème, mais ne se base plus sur les mêmes règles de connaissances intrinsèques de la conception du monde des chercheurs en IA. L'IA connaît une ouverture sur le monde extérieur. La finalité de ces programmes est d'émettre des réponses à des questions, soit résoudre des problèmes sous la forme d'exemple et de contre-exemple en effectuant un raisonnement à partir de règles établies par l'expertise envers divers domaines. On cherche à travers ces systèmes à mettre en place des algorithmes répondant à la demande des investisseurs ; soit des dispositifs en phase avec le monde extérieur en prenant appui sur la confiance accordée au savoir expert (cf chapitre 1 -2.2.1.) Si l'on ne peut pas réellement parler d'une nouvelle définition de l'intelligence, le système expert fait toutefois figure de transition dans la représentation du concept.

### 2.1.3. L'optimisation

La métaphore neuronale fait son retour avec l'apparition du courant connexionniste à travers laquelle se dessine une nouvelle définition de l'intelligence, ou du moins la fin de l'intelligence en tant que résolution de problème. Contrairement à l'idée reçue, la recherche en IA n'a pas fait de découvertes théoriques récentes et révolutionnaires, elle a seulement repris une méthode qui existait déjà dans un contexte qui lui permet de fonctionner. Notre utilisation

d'Internet a bouleversé notre quotidien de manière si profonde et globale que la prolifération massive de données résultant de ce phénomène est appréhendée aujourd'hui comme la représentation la plus proche, précise et réaliste de notre monde. Cette appréhension du monde, bien que largement questionnable, permet selon les chercheurs connexionnistes de franchir la limite insurmontable présentée par le philosophe Dreyfus dans sa critique envers l'IA symbolique selon laquelle : « le meilleur modèle du monde est le monde lui-même ». Or, les sociologues et auteurs de l'article « La revanche des neurones » montrent bien que ce modèle du monde se base sur des données dont la nature leur a été dépouillée :

« Alors que les mondes « jouets » et les mondes de connaissances des *machines symboliques* étaient faits de petits univers restreints, nettoyés et domestiqués par une grille de caractéristiques intelligibles et interdépendantes, les *machines connexionnistes* se déploient dans un monde dans lequel les données doivent non seulement être massives, mais aussi les plus atomisées possible afin de leur ôter toute structure explicite. » (Benbouzid et Cardon, 2018 : 288)

Le phénomène du BigData s'est substitué à toutes tentatives de modélisation humaine fondée sur un courant de pensée théorique précis. Ainsi la valeur d'un système d'IA se trouve dans les résultats techniques et non plus désormais dans la réussite à modéliser le raisonnement humain au plus proche de sa configuration biologique. Le connexionnisme n'est né d'aucun courant de pensée théorique. Si aujourd'hui des milliards sont dépensés dans la recherche de l'apprentissage profond, ce n'est donc pas parce qu'on a enfin décrypté les secrets de l'esprit humain, mais simplement parce que les méthodes empiriques, accumulées à la quantité de données, la capacité de stockage et la puissance de calculs des ordinateurs permettent des résultats spectaculaires. Par résultats spectaculaires, les chercheurs entendent une capacité à percevoir et prédire des caractéristiques au sein d'une masse gigantesque de signaux sensibles. Autrement dit, la sortie voulue du système d'IA ne caractérise plus vraiment la représentation de l'intelligence en IA puisque l'ambition de base idéologique disparaît progressivement. Aujourd'hui, la fonction première de l'apprentissage profond reste descriptive. Elle permet de rendre compte d'un fait ou d'un phénomène déjà observé et de confirmer l'observation par le modèle (pattern) établi dans la production de données. C'est l'apprentissage supervisé qui est alors utilisé : le système classe selon un modèle de classement préalablement établi. L'être humain reste actif dans la prise de décision puisque c'est lui qui fait état de l'observation et qui interprète les résultats du programme. (Merveille, 2018) Nous entrons peu à peu dans la fonctionnalité du prédictif. De grandes entreprises utilisent déjà les applications de l'apprentissage automatique dans cette visée à des fins marketing. Les modèles établis à partir des données de consommation de leur clientèle permettent de prédire les ventes et d'optimiser

ainsi la production. C'est l'apprentissage semi-supervisé qui est alors utilisé. C'est aussi la méthode prescrite pour l'aide au diagnostic médical. On sait globalement ce qu'on cherche, mais on laisse à l'algorithme la liberté de détecter les spécificités. Le programme quantifie le niveau de probation, sur lequel l'individu va se baser pour établir sa stratégie d'action. L'Homme commence donc ici à être influencé dans sa prise de décision.(Merveille, 2018) L'idée est d'éventuellement utiliser les applications de l'apprentissage profond pour sa fonction prescriptive. L'algorithme formulerait alors la réponse à émettre. En utilisant un apprentissage non supervisé, cette fois, on laisse à l'algorithme l'autonomie d'établir seul des connexions, des modèles et de produire de l'information. C'est le système qui découvre par lui-même la structure. L'Homme ne participe plus à la prise de décision. En pratique, cette fonctionnalité est encore très peu exploitée. Ce n'est pas parce que l'IA n'est pas rendue là, mais plutôt parce que l'Homme n'est pas prêt à se laisser complètement guider aveuglement par la machine d'après Nicolas Merveille, anthropologue spécialiste de l'Internet des objets. (Merveille, 2018).

L'optimisation constitue donc globalement la sortie recherchée des programmes développés aujourd'hui en IA. Par l'intermédiaire du calcul statistique, la visée est de minimiser ou maximiser un comportement afin d'en soutirer le meilleur résultat possible selon l'objectif. Autrement dit, il s'agit d'émettre un rendement efficient sous le critère de l'optimisation. L'élément à la sortie du système doit dépasser en valeur l'élément en entrée en matière de performance. À l'image d'un retour d'investissement, les données de bases intégrées par le programme d'apprentissage profond seront *optimisées* à la sortie.

Certains s'efforcent encore de fournir la preuve d'une plausibilité biologique connexe à la fonction mathématique efficace pour réduire la fonction d'erreur au sein d'un programme informatique (O'Reilly 1996). Il est toutefois évident que l'on s'éloigne progressivement d'une véritable volonté idéologique. Alors que les chercheurs de l'IA symbolique prenaient le temps de théoriser leur recherche empirique et de justifier leur théorie en revenant toujours à leur conception du modèle biologique humain, il suffit de lire les articles publiés par Yoshua Bengio (ou tous autres comptes rendus des recherches contemporaines en IA) pour se rendre compte que l'ambition s'est déplacée (Bengio, Courville et Vincent, 2012).

Le fait que l'on s'éloigne progressivement du projet idéologique peut hypothétiquement révéler un certain pessimisme quant à la possibilité de créer une intelligence artificielle à l'image de l'Homme tant le projet est vain. Cela voudrait dire que les chercheurs savent que l'intelligence humaine n'est pas réductible à une simple modélisation informatique. Ils préféreraient donc concentrer leur recherche sur les aspects techniques et les applications

pratiques qui peuvent découler de cette poursuite idéologique infructueuse. Cependant, la fin de cette poursuite implique une rupture définitive avec les fondements humanistes. En effet, la volonté de se saisir des mystères de notre espèce à travers le projet de créer une intelligence artificielle (certes par fascination anthropocentrique) entretenait le dernier lien restant avec l'humanisme des Lumières. Cette fracture se traduit par une dévalorisation de la recherche fondamentale comme quête de la vérité (idéologie des Lumières), au profit de la recherche opérationnelle comme quête d'efficacité (idéologie de la société postmoderne). La transformation s'effectue aussi bien au niveau de la forme de la recherche (méthodes qui ne sont plus bio-inspirées), des objets de l'enseignement, que dans la finalité (applications technologiques au détriment des théories sur l'esprit humain). La visée se base sur les besoins de la société de consommation et d'innovation. Si les intérêts économiques et professionnels existaient déjà au temps de l'IA symbolique majoritairement financé par l'aile militaire, ce n'était pas la finalité première recherchée ; aujourd'hui, oui. La logique de l'efficacité et du profit prime sur la logique de la raison et de la simple acquisition du savoir. L'IA symbolique tentait encore de mêler la connaissance pure synthétique et proprement théorique au caractère essentiellement instrumental de la méthode empirique. L'IA aujourd'hui ne focalise la recherche que sur la pratique, délaissant à jamais tout fondement conceptuel. Dorénavant, seul ce qui sert les intérêts corporatifs est valorisé et subventionné. Sous le prétexte d'un développement technologique pour le bien social (à travers le développement d'applications s'attaquant à l'aliénation du travailleur, à la crise environnementale, à la famine, etc.) la conquête de l'IA par les entreprises et les nations poursuit la quête à la puissance économique.

Peut-on alors parler d'une nouvelle représentation de l'intelligence lorsque l'humain n'est plus au centre des considérations de cette discipline qui se dit « scientifique » ? D'une certaine manière, nous pensons que oui. En effet, on retrouve dans le critère de l'optimisation une caractéristique si valorisée dans nos sociétés néolibérales consuméristes et marquetées qu'elle rentre intrinsèquement dans la représentation de l'Homme ou plutôt de ce qu'on attend de lui. Selon ce postulat, ce n'est plus la rétroaction qui définit l'intelligence, ni même la capacité d'un individu de résoudre des problèmes. Quelqu'un d'intelligent serait quelqu'un dont le rendement est efficient. C'est celui qui sait optimiser les ressources dont il dispose afin d'émettre les meilleures solutions possible, meilleur dans le sens qui sert un intérêt de productivité de temps ou d'argent. L'intelligence se définirait donc dorénavant en terme instrumental dans une logique de production et de contrôle.

L'IA, voulant reproduire au plus près ce qui est considéré être l'intelligence humaine, s'est adaptée spontanément aux qualités et normes les plus valorisées (et donc constituant l'intelligence) d'une société en constante mouvance. C'était tout d'abord la capacité de rétroagir avec son environnement, puis la capacité de résoudre des problèmes complexes et, finalement, le potentiel d'optimisation sur objectif.

## **2.2. Une représentation réductrice de l'Homme**

L'objectif de cette démarche (soulever les représentations de l'intelligence inscrite en IA) n'est pas de remettre en question le fondement scientifique des mécanismes cognitifs et neuronaux. Il s'agit plutôt de démontrer que l'IA repose sur un certain réductionnisme qui engendre de réels enjeux épistémologiques. En effet, comme nous l'avons vu, l'être humain et son intelligence doivent être réduits à de simples mécanismes biologiques pour que le projet de l'IA soit réalisable. Les théories et approches développées à partir des années 50 et sur lesquels se fonde l'IA prennent vie au sein du même cadre informationnel. Ce paradigme préconisant une approche informationnelle pour l'étude de toutes entités et phénomènes, les mécanismes biologiques sont donc eux-mêmes réduits à des processus modélisés et donc formalisables dans un langage que la machine peut facilement manipuler. Pour récapituler ce qui a été vu à la partie 2.1 (Une définition en évolution), l'intégralité de l'intelligence humaine se résume en IA par la capacité d'analyse et d'observation, l'exécution d'opérations par la manipulation de chiffres et la compréhension de phénomènes complexes en faisant preuve de logique opératoire. Dans cette représentation logico-mathématique de l'être humain, ces facultés sont considérées comme substantielles à l'Homme. L'approche réductionniste empruntée par les chercheurs en IA leur permet d'affirmer qu'il est possible de transmettre l'intelligence à la machine.

On discerne toutefois une certaine volonté de s'extraire des définitions de l'intelligence (rétroaction, résolution de problèmes, expertise, rendement efficient) enfermées dans le paradigme informationnel. La théorie des intelligences multiples de Gardner, professeur en cognition et éducation, est un des exemples les plus démonstratifs. Il évoque lui-même cette intention dans *Les intelligences multiples* : « De ce fait, c'est à ce niveau que la théorie des intelligences multiples rompt avec la tradition, qui définit l'intelligence d'une manière opératoire, comme l'aptitude à répondre aux items des tests d'intelligence. »(Gardner, 2009 : 37) L'utilisation du terme « tradition » souligne la volonté de sortir du cadre établi et de la

représentation dominante empruntée par l'IA. Il décide malgré tout de conserver le terme « intelligence » pour parler des différentes facettes de l'être humain. Il suggère que l'intelligence n'est pas singulière, mais multiple. En effet, pour lui, les êtres humains seraient doués de sept intelligences : l'intelligence musicale, kinesthésique, logico-mathématique, langagière, spatiale, interpersonnelle et intrapersonnelle. À l'encontre des chercheurs Mugny et Carugati, le théoricien des intelligences multiples postule que l'intelligence fait partie de notre héritage génétique. Au niveau élémentaire, elle se manifesterait de façon universelle, quelle que soit la socialisation de l'individu. L'intelligence serait, d'après lui, un « potentiel biopsychologique » (Gardner, 2009 : 49). Malgré la tentative de développer un nouveau modèle de pensée, sa théorie reste donc profondément ancrée dans ce qu'il appelle « la tradition » et ce que nous appellerons le modèle informationnel. La question n'est pas de savoir si sa conception est recevable. Cette théorie souligne simplement une hypothèse pertinente pour notre propos : l'intelligence telle qu'elle est comprise au sein du paradigme informationnel ne prendrait en compte qu'une mince partie de la complexité de l'être humain. L'IA serait alors la reproductibilité de cette forme d'intelligence, que Howard Gardner nommerait « intelligence logico-mathématique » (Gardner, 2009 : 41) ; une intelligence opératoire et donc mesurable sans prise en considération des autres conceptions de l'intelligence, ou autres facettes de l'être humain.

L'étude de Darsen (développé au point 1.1 : À la lumière des représentations sociales) permet aussi de rendre compte du réductionnisme de la conception de l'intelligence adopté par l'IA. En effet, la représentation de l'intelligence de la population des Baoulés met en lumière un trait de l'être humain qui n'est pas inclus comme priorité de l'intelligence en IA : le sens de la communauté. Cet attribut est lié de près au concept de l'affect, car il fait appel à l'empathie et aux émotions. Même si ce qualificatif paraît constituer un élément essentiel pour la fabrication d'une créature à l'image de l'Homme, les machines n'auraient pas besoin d'émotions pour devenir intelligentes au sens attribué à cette notion en IA. Interrogé par *Le Monde* sur la question des émotions de la machine, le philosophe Jean-Michel Besnier tente d'expliquer pourquoi :

« Ce sont les signaux qui comptent. En toute rigueur, votre intériorité, je ne sais pas si elle existe... Elle se manifeste à moi par des signes extérieurs. Je ne sais pas ce que c'est que de ressentir une émotion chez un autre être humain. Je peux ressentir de l'empathie, mais je peux aussi en ressentir face à un acteur, qui simule. Je ne peux pas savoir si ces signes émanent d'une intériorité ou d'une simulation. C'est pourquoi pour ceux qui fabriquent des machines, si elles sont capables de simuler, ça suffira. »(Tual, 2015)

Les spécialistes en IA ne prétendent pas que les émotions n'interviennent pas dans la prise de décision de l'individu, mais suggèrent qu'elles ne sont pas absolument nécessaires pour parvenir à « des résultats tangibles et démontrables. » (Ganascia, 2007 : 75) Afin d'estimer les avancées en IA, les chercheurs comparent sur une même échelle « l'intelligence » du système et l'intelligence humaine. Ils regardent ce que la machine est capable d'exécuter et comparent avec les capacités humaines. L'intelligence est ainsi évaluée par le résultat, l'aboutissement et non par le processus sous-jacent (Gardner, Kornhaber et Wake, 1996 : 165). Pour illustrer ce point, il est pertinent de se pencher brièvement sur la question de l'art en IA. Les tableaux créés par les systèmes d'IA sont jugés par beaucoup d'expert en histoire de l'art comme de même qualité que l'œuvre des plus grands artistes de notre temps. Mais peut-on réellement juger de la valeur d'une pièce d'art seulement par l'aboutissement, l'objet de la toile sans prendre en compte le processus sous-jacent ? Le sujet, et donc l'artiste, ne fait-il pas partie intégrante de l'œuvre ? Peut-on dissocier le processus de création de l'artiste de son inspiration guidée par ses émotions ? N'étaient-ce pas ces mêmes émotions qu'il retranscrit sur la toile ? Le résultat peut sembler tout aussi bon, mais l'IA n'est pas celle à l'origine de l'idée, elle n'est pas douée de créativité parce qu'elle n'est pas affectée par une histoire personnelle. Autrement dit, le système d'IA connaît les codes, mais ne sait rien de leur signification.

Le philosophe John Searle étudie cette question de la signification des symboles à travers une critique de l'IA. Un programme d'IA serait capable d'intégrer de l'information en langage formel et de résoudre le problème auquel il est confronté. Le système computationnel va assimiler le processus et mettre en place une stratégie répondante. Mais est-ce que ça veut dire qu'il a réellement « compris » la signification de l'information ? Selon le philosophe, la réponse est négative.

« Because the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even symbol manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have only a syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output. » (Searle, 1980 : 11)

Il explique que la sémantique n'est pas importante dans les programmes informatiques, seule la syntaxe est prise en compte. Or, l'intentionnalité se trouve selon lui dans la signification des symboles et donc dans la sémantique. La machine n'a pas d'intentionnalité, car elle ne fait que manipuler des symboles dont elle ne saisit pas le sens. Dans l'article « Mind, Brains and Programs » (Searle, 1980) publié en 1980, il tente de démontrer que l'intentionnalité est le

vecteur qui prouve que la reproductibilité du comportement humain n'est pas possible sous forme d'un programme informatique. L'IA se base sur le postulat selon lequel l'intentionnalité est un produit des caractéristiques causales du cerveau. Autrement dit, l'intentionnalité ne serait que le résultat de certains processus cérébraux. Searle réfute ce postulat. Pour lui, elle est certes un produit du cerveau, mais elle est aussi indissociable du vécu et de la conscience du sujet dû à sa confrontation au monde extérieur. L'intentionnalité serait donc, selon cette vision, la capacité biologique de l'esprit de mettre l'organisme en rapport avec le monde environnant tout en étant conscient de cette expérience subjective. Contrairement à la généralisation de type scientifique, Searle croit fermement que l'intelligence n'est pas conceptualisable en raison du caractère privé des états intentionnels tels que les croyances et les désirs qui, eux, ne peuvent être généralisés (Andrieu, Anderson, Le Parvis, 1992 : 25). Le fait d'exemplifier un programme informatique ne revient donc pas à donner de l'intentionnalité à la machine. C'est l'impossibilité de reproductibilité de l'intentionnalité, propre aux êtres vivants, qui démontre, selon lui, que le projet d'IA (forte) est un projet vain.

La critique de Dreyfus, considéré comme la plus féroce de son temps, remet aussi en question la définition de l'intelligence prise pour acquise en IA symbolique en mettant en avant l'irréductibilité de l'être humain face à la machine. Herbert Dreyfus est professeur de philosophie au MIT, la même université où le pionnier Minsky conduit ses recherches en IA. Confronté de près à cette nouvelle mouvance donc, Dreyfus va rapidement s'intéresser à la question de « la machine pensante » d'un angle philosophique. Il remet en question les présupposés sur lesquels s'appuient les recherches du domaine qui leur permettent d'affirmer que l'intelligence est reproductible à la machine dans son livre *Alchemy and Artificial Intelligence* (Dreyfus, 1965). Selon lui, quatre présupposés prédominent : le présupposé biologique, psychologique, épistémologique et ontologique. Les deux derniers présupposés sont ceux qui nous intéressent et sur lesquels Dreyfus s'appuie pour fonder sa théorie critique. Le présupposé épistémologique consiste à penser que toute action peut être formalisée mathématiquement selon un système de lois prescrites. Sur cette base, les experts peuvent affirmer que l'intelligence se retranscrit comme activités mentales provenant d'un système de règles. Quant au présupposé ontologique, il consiste à considérer la réalité comme un ensemble de faits indépendants les uns des autres. Les connaissances d'un individu seraient donc, selon ce présupposé, un de ces faits intrinsèques de la réalité. Autrement dit, selon les représentations des chercheurs en IA symbolique, l'individu répondrait conformément à des lois prescrites et fixes dans le temps. Herbert Dreyfus, lui, défend le fait que l'individu « sait faire » sans savoir

expliquer le processus qui précède l'action et le comportement. C'est précisément ce fait qui prouve selon lui que *le savoir-faire* (« know-how ») ne nous est pas accessible à travers un ensemble de règles et de faits prescrits d'avance, mais à travers l'expérience et la pratique.

«You probably know how to ride a bicycle. Does that mean you can formulate specific rules that would successfully teach someone else to do it ? How would you explain the difference between the feeling of falling over and the perfectly normal sense of being slightly off balance when turning ? And do you really know, until it happens, just what you would do in response to a certain unbalanced feeling ? No, you don't. You can ride a bicycle because you possess something called « know-how » which you acquired from practice and sometimes painful experience. The fact that you can't put what you have learned into words means that know-how is not accessible to you in the form of facts and rules. If it were, we would say that you « know that » certain rules produce proficient bicycle riding.» (Dreyfus, Dreyfus et Athanasiou, 1986 : 16)

Si les lois de la physique peuvent expliquer ce qui se passe lorsqu'on tombe à vélo, la différence lorsqu'on tourne et lorsqu'on se retrouve en déséquilibre, aucune loi ne peut expliquer la sensation que ressent l'individu. Celui-ci arrive pourtant à faire la différence et à agir en conséquence même s'il ne saurait l'expliquer. Pour le philosophe, l'individu peut exécuter une action sans pour autant savoir expliquer le processus physique engendré durant cette action. Le sens commun dont l'Homme se sert pour la plupart des actions qui guident son quotidien n'est pas accessible à la machine puisqu'il ne peut être retranscrit sous forme de règles prescrites, mais à travers l'expérience.

De plus, selon ces présupposés l'individu répondrait à un ensemble de règles prescrites ; le comportement humain serait donc indépendant du contexte. Cet angle de perspective permet de légitimer leur méthode : l'IA décrypte les règles internes de l'esprit et du comportement humain tout comme les lois de la physique décryptent le monde extérieur naturel. C'est ce postulat que rejette Dreyfus, conformément à la tradition phénoménologique. Selon lui, notre être est indissociable du contexte global et la perspective cognitiviste empruntée par les chercheurs en IA ne permet pas de prendre en compte ce contexte. En réduisant l'humain à un ensemble de lois scientifiques, ils objectivent selon le philosophe le comportement humain. L'IA est donc basée sur l'hypothèse que la pensée peut être analysée comme un ensemble fini d'opérations simples déterminées. Selon Dreyfus, la pensée impliquerait plutôt trois sortes de processus globaux qui, ensemble, forment un tout (Nilsson, 2010 : 312). Le premier est ce qu'il nomme *la conscience de frange* (« fringe consciousness ») qui représente tout ce que le cerveau utilise pour avoir accès à l'information extérieure du quotidien. Cette conscience permet de considérer à la fois les détails et la vue d'ensemble. Le second processus global permettrait de distinguer l'essentiel de l'inessentiel. Enfin, le troisième est le contexte global. Combinés, ces

trois processus permettent au cerveau de reconnaître des modèles complexes comme les visages humains : c'est ce qu'il appelle *le groupement perspicace*, en anglais « *perspicuous grouping* ». Selon lui, la machine, et plus spécifiquement l'ordinateur, n'a pas la capacité d'appliquer ces processus globaux propre au cerveau humain et pourtant essentiel au comportement intelligent (Nilsson, 2010 : 312).

En s'attaquant à la conception d'une machine pensante à l'image de l'Homme, les chercheurs en IA prétendent avoir une idée suffisamment claire du fonctionnement du raisonnement humain. Or, cette question a attisé l'intérêt des philosophes depuis plus de deux siècles sans réponses concrètes et définitives. La différence entre les théories philosophiques qui s'attèlent à décrypter l'esprit humain depuis plus de deux siècles et la discipline « scientifique » que représente l'IA, c'est que les premiers ne prennent pas des possibilités de représentations pour des principes absolus. Ce que Herbert Dreyfus veut réellement nous inculquer, c'est que la question de l'esprit humain est une question complexe à laquelle nous n'aurons probablement jamais de réponses définitives absolue et exacte et « qu'il en soit ainsi ». L'IA pense pouvoir reproduire l'intelligence à travers la machine en se basant seulement sur une théorie parmi d'autres. Selon Dreyfus, pour que le projet idéologique de l'IA se concrétise, il faudrait pouvoir créer un modèle intégralement similaire à notre être. Il faudrait donc créer un modèle d'un corps similaire au nôtre. « That is, we would have to include in our program a model of a body very much like ours with our needs, desires, pleasures, pains, ways of moving, cultural background, etc. » (Dreyfus, 2007 : 247) Or, ceci est selon le philosophe, tout simplement impensable en pratique.

Théodore Roszak émet ce même constat à travers une critique du paradigme informationnel qu'il nomme « culte de l'information ». La notion de l'information est empruntée pour tout et vénérée de tous sans intérêt et réelle compréhension des fondements de ce concept : d'où l'appellation de « culte ». Pour l'historien, il y a une distinction vitale entre le traitement de l'information et la pensée qu'il convient de saisir et de conserver. Au sein de ce modèle informationnel, nous l'avons vu, la distinction ontologique entre l'être humain et la machine est inexistante en nature.

« Because the ability [of the computer ] to store data somewhat corresponds to what we call memory in human beings, and because the ability to follow logical procedures somewhat correspond to what we call reasoning in human beings, many members of the cult [of information] have concluded that what computers do somewhat correspond to what we call thinking. » (Roszak, 1986 : x)

On comprend ici que, selon l'historien, cette distinction serait la frontière qui nous préserve du piège à croire que l'intelligence humaine correspond à un simple traitement de l'information et donc, que celle-ci est transférable à la machine. Pour lui, le fantasme d'une IA forte n'est pas inquiétant puisqu' irréalizable. « There is no possibility that computer will ever equal or replace the mind except in those limited functional applications that do involve data processing and procedural thinking. »(Roszak, 1986 : 132) Ce dont il se méfie, en revanche, c'est de l'utilisation inappropriée des personnes convaincues. Les programmes d'IA pourraient alors être utilisés pour des tâches qui demandent une réflexion qui va au-delà d'un simple traitement de l'information.

Au vu des nombreuses caractéristiques intégrantes à l'être humain qui ne sont pas prises en compte dans la représentation de l'intelligence en IA, il est raisonnable d'affirmer que l'IA prône une conception extrêmement réductrice de l'être humain et de son intelligence.

### **3. Artificialisation de l'intelligence humaine et naturalisation de l'IA...**

#### **3.1. ... à travers le reversement du dualisme nature-artifice.**

Les chercheurs s'inspirent du fonctionnement de la nature humaine pour créer une machine, tout en s'inspirant du fonctionnement de la machine pour comprendre la nature humaine. Ce fait paradoxal renvoie au double processus dont fait état Daphné Esquivel Sada dans son mémoire *Le « nanomonde » et le renversement de la distinction entre nature et technique : entre l'artificialisation de la nature et la naturalisation de la technique* : celui de « l'artificialisation de la nature », technique qui imite la nature, possible si cette dernière est appréhendée comme une machine et de « la naturalisation de la technique », progrès technologique perçu comme un fait naturel inéluctable (Esquivel Sada, 2009). Ce double processus illustre parfaitement le brouillage des frontières entre nature et technique sans lequel les deux termes « intelligence » et « artificielle », aussi antagonistes qu'ils puissent paraître, ont pu être rassemblés en un seul et même concept, donnant lieu à une discipline majeure au 21<sup>e</sup> siècle : « l'Intelligence Artificielle ». Daphne Esquivel Sada tente de démontrer en quoi le rapport nature et artifice relève d'un enjeu majeur dans notre société, aussi bien dans la construction du dualisme propre aux sociétés occidentales modernes, que dans sa remise en question. Elle prend pour objet de recherche les nanotechnologies. Il paraît clair cependant, à

la lecture de ce mémoire, que l'enjeu se s'arrête pas à l'échelle du nanomonde ; ce champ technoscientifique spécifique permet simplement d'illustrer particulièrement bien ce phénomène. En s'appuyant sur ce double processus, on peut tenter d'expliquer en quoi l'IA s'est établie à partir du basculement du dualisme nature/ technique.

L'intelligence est comprise depuis Darwin comme une entité biologique reliée à l'organe du cerveau ; dans cette perspective elle appartient à la catégorie du naturel. Les termes choisis « intelligence » et « artificielle » pour représenter ce champ technoscientifique en disent long sur la distinction nature/ technique. Il est bien question de rendre artificiel ce qui est naturel. En ce sens, le processus d'artificialisation de la nature ne pourrait être mieux représenté que par le champ de l'IA. Esquivel Sada rappelle que les prémices du phénomène d'artificialisation de la nature remontent aux débuts de l'agriculture, mais illustre une véritable dissolution des frontières qu'à partir de l'avènement du paradigme cybernétique (Esquivel Sada, 2009 : 99). Rappelons que ce nouveau champ scientifique dominant a permis de reformuler les grands principes de la recherche scientifique et, dans une société érigée par la science, a ouvert les portes à une redéfinition complète de nos catégories. La théorie unificatrice découlant de la cybernétique s'applique aussi bien pour l'humain que pour le non-humain. Le monde naturel autant que le monde artificiel s'explique par le biais de la théorie de l'information et du concept de la rétroaction. L'introduction par la cybernétique du renversement du dualisme nature et technique légitime le phénomène d'artificialisation de la nature. Au sein du modèle informationnel, toute matière naturelle est perçue et comprise comme un système mécanique pouvant être traduit en langage formel binaire. Ce n'est plus seulement en partant de la compréhension de l'être humain dans une représentation spécifique que l'on conçoit la machine, mais à partir du fonctionnement de la machine que l'on explique le fonctionnement organique de l'Homme. L'intelligence est ainsi appréhendée comme un programme informatique. Sachant que toute entité naturelle comme technique fonctionne selon le même mécanisme informationnel, il est possible de considérer que tout phénomène de la nature vivante, comme l'intelligence, peut-être simuler au sein d'une machine.

Le processus d'artificialisation de la nature est légitimé, entre autres, par la théorie de l'acteur réseau développé par le sociologue, anthropologue et philosophe Bruno Latour dans les années 1980. Selon cette théorie, s'apparentant au courant du nouveau matérialisme, le monde social se constitue à travers l'interaction de toute sorte de matière sous la forme de réseaux (Latour, 2010). Dans cette conception, la matière non humaine a un rôle actant au même titre que l'humain. Le dualisme propre à la modernité ne serait en fait que discursif. Selon Latour, l'humain et le non-humain (animal, plante, molécule, objet, discours, etc.) sont pris

sous une fausse dichotomie qui se doit d'être rassemblée. Sous l'égide théorique d'un dualisme permettant de réduire la complexité du réel ; en pratique, ce dualisme discursif aurait permis à la science de complexifier encore plus le réel. Plutôt que de considérer le monde à travers cette barrière infranchissable qui sépare ce qui est propre à la nature, et ce qui est propre à l'Homme (la culture, la technique, l'artifice), on devrait, selon Latour, le considérer à travers un continuum horizontal entre nature et artifice si l'on veut conceptualiser le monde dans lequel on évolue de manière plus juste. Cet enchevêtrement de catégories, cette tendance à l'horizontalisme dans l'indifférenciation de l'agentivité, légitime l'artificialisation de la nature, soit l'appropriation technique du monde naturel, et donc ici, l'appropriation technique de certaines fonctions cognitives comprises comme constituantes de l'intelligence humaine.

Quant au processus de « naturalisation de la technique », il suggère l'idéologie selon laquelle la technique est intimement associée aux faits naturels. Esquivel Sada illustre ce processus par le concept d'autonomisation du développement du nanomonde ; développement autonome dans le sens qu'il existe par lui-même et serait donc humainement non maîtrisable. Cette idée renvoie à la conception d'un progrès technoscientifique qui s'inscrit dans une interprétation faussée de la théorie de l'évolution de Darwin, si l'on se fie à Pascal Picq dans l'article « « Homo » et la fin des certitudes » (Latour, 2010). Cette interprétation affirme que selon le principe de sélection naturelle, tout ce qui serait bénéfique à l'Homme et à la société perdurerait inévitablement au détriment de ce qui serait le moins souhaitable et qui ne participerait pas à la survie de l'espèce humaine. « Autrement dit, ce qui est compris comme une finalité s'impose comme un processus pétri de certitudes : s'il en a été ainsi, c'est qu'il devait en être ainsi ». Il ajoute, « Il n'existerait donc qu'une seule voie achevée de l'évolution. » (Picq, 2014 : 23)

Cette même conception se retranscrit dans le projet de l'IA. Le stade de l'intelligence transmis à la machine semblerait aller dans le sens du progrès technoscientifique, un progrès prétendument bénéfique, puisqu'inéluctable. L'IA s'entend, selon le processus de « naturalisation de la technique », comme une avancée technologique qui va au-delà de la maîtrise de l'Homme, comme une avancée qui suit le cours naturel de l'évolution de l'humanité. Bernard Andrieu, dans *Les intelligences. Animal-Homme-Machine*, déconstruit le présupposé selon lequel l'IA est un stade de l'évolution au même titre que la bipédie. Selon lui, l'intelligence animale s'inscrit dans la logique de la sélection naturelle au sens qu'elle permet l'adaptation au milieu sans transformation par la technique, contrairement à l'intelligence humaine qui adapte son milieu à ses besoins par le moyen de la technique. « L'évolution des

espèces est désormais brisée par l'Homme : en effet sa technique bouleverse l'écosystème au point de le déséquilibrer définitivement. » (Andrieu, Anderson, Le Parvis, 1992 : 99) Selon cette logique, il n'est désormais plus possible d'expliquer les phénomènes humains par la théorie de l'évolution. Par la technique, l'humain aurait déjoué la simple logique de sélection naturelle. Dans cette perspective, l'IA en tant que technoscience ne peut être considérée comme phénomène naturel faisant partie de la linéarité inéluctable de l'évolution de la vie.

Cette perspective appartenant au constructivisme social s'oppose à la perspective adoptée par les chercheurs en IA et représentative du modèle informationnel : le déterminisme technologique. Le progrès technique, selon cette approche, est autonome et donc indépendant des effets de la société. Le déterminisme technologique conduit à considérer l'IA comme une avancée naturelle inévitable née des mêmes lois que tous phénomènes et composants de ce monde. Compris comme phénomène naturel, le progrès technique provoque un changement social incontournable dont la direction dépendrait intégralement de la discussion envers les enjeux éthiques, économiques et sociaux. Le déterminisme technologique est omniprésent dans les discours des chercheurs en IA : on ne peut arrêter le cours naturel des choses et l'avancée du progrès technologique amenant à la création prochaine d'une IA, mais nous pouvons réfléchir aux solutions pour contrer les effets négatifs que cela va inévitablement provoquer sur la société. En se limitant à débattre des enjeux de l'IA, dont le seul but est de minimiser les risques et maximiser les avantages, le fondement même de l'IA est rarement contesté (Le Dévédec, 2015b : 4). Le projet de l'IA est ainsi légitimé par le processus de naturalisation de la technique.

On retrouve ce processus dans les approches d'avenir à la fois fatalistes et optimistes qu'on retrouve parfois dans la littérature contemporaine. Selon ces perspectives, il serait vain de nier que la technique est partie intégrante de notre univers. Le monde naturel se construirait à présent autour et en relation avec la technique, si bien que ces deux catégories sont maintenant indissociables. Cette idée fait écho à la publication de Kim de Wolff « Plastic Natureculture » (De Wolff, 2017). Elle utilise l'exemple de la matière plastique afin d'illustrer l'impossibilité de séparation de la technique et de la nature de nos jours, et donc dans l'article, du plastique et du vivant. Le danger ne se situe pas, selon elle, dans la présence du plastique dans les océans puisque la nature sait s'accoutumer des changements causés par l'être humain, contrairement à ce que prétend la vision anthropocentrisme. Le réel danger, selon l'auteur, se trouve justement dans la séparation de cette matière plastique et de la manière vivante. Elle argumente que certaines espèces marines se sont construites en harmonie avec la matière plastique. L'environnement naturel de ces espèces est fait de plastique ; le plastique est donc leur nouvelle

nature. On retrouve très clairement le processus de naturalisation de la technique dans cette perspective. La nature s'approprié la technique, la fait sienne, indépendamment de l'Homme. La technique serait alors indissociable du monde naturel et inversement.

Le processus de naturalisation de la technique, lui aussi légitime le développement de l'IA en considérant cette technoscience comme une forme d'intelligence sur la même lignée de l'évolution de la vie humaine. L'indifférenciation entre ce qui est naturel et technique permet de mettre sur le même plan et simplement à différents degrés l'intelligence humaine et le programme informatique simulant une représentation spécifique de l'intelligence.

Daphne Esquivel Sada démontre que ces deux processus, bien que paradoxaux en apparence, fonctionnent en cohésion. L'hybridation de la nature et de la technique dépend d'ailleurs de l'enchevêtrement de ces deux processus. Pour résumer grossièrement, l'imitation de la nature dans l'élaboration de la technique est justifiée par le caractère naturel donné à la technique. Un parallèle avec la théorie sortante de ce mémoire et notre objet de recherche suggère le postulat selon lequel le projet d'une intelligence artificielle n'est plausible que par la déconstruction du dualisme nature/ technique.

### **3.2. ... à travers l'idéologie transhumaniste**

On retrouve les deux processus illustrés ci-dessus dans les discours et projets des transhumanistes qui reprennent à travers leur idéologie la conception instrumentale et opérationnelle de l'intelligence.

Le transhumanisme est un courant de pensée née aux États-Unis dans les années 1980. Ce mouvement intellectuel prône l'utilisation des technosciences dans une visée d'amélioration de la condition humaine. L'augmentation des facultés biologiques (aussi bien physiques que mentales) par les moyens de la technique est de vigueur. Ce mouvement, porté par quelques futurologues, ingénieurs, philosophes et autres ne compte qu'une poignée de membres dans le monde. En revanche, ces quelques adhérents se trouvent être des figures d'autorité et jouissent ainsi d'une influence considérable. Industrie, recherche fondamentale et idéologie se mêlent à travers ce mouvement, lui conférant une légitimité non négligeable. Ray Kurzweil est *la* figure emblématique du transhumanisme. Mathématicien et futurologue, auteur de nombreux ouvrages faisant la promotion de l'idéologie transhumaniste, parfois critiqué pour ces discours extrêmes et prédictions démesurées, il suscite autour de lui une fascination presque

« religieuse ». Pour preuve, la presse le surnomme « le pape » du transhumanisme. Maintenant ingénieur principal chez Google (Le Dévédec, 2015a : p.198), son influence est d'autant plus importante qu'elle est soutenue par une des plus grandes multinationales ayant une capacité d'investissement suffisante à la recherche et au développement d'applications concrètes basées sur les principes transhumanistes.

La finalité transhumaniste n'est plus seulement une meilleure compréhension des processus cognitifs, mais le dépassement des compétences humaines afin de les perfectionner en performance automatique. Pour les transhumanistes, la fusion de l'IA avec l'être humain représenterait la seule et unique solution de survie à l'être humain. L'Homme se serait créé un environnement si technicisé, qu'il devrait à présent s'y adapter s'il veut vivre en son sein ; se transformer soi-même, pour être à la hauteur de la technique qu'il a créée. Voici l'idée du « modèle posthumaniste de la perfectibilité » (Le Dévédec, 2015a : p.197) soutenu par les transhumanistes et déconstruit par le sociologue Nicolas Le Dévédec à travers la thèse *La société de l'amélioration* (Le Dévédec, 2015a). On entend par là une sollicitation extrême : il ne sert à rien d'essayer de contrer le cours « naturel » de la technique dont le projet de création d'une IA est partie intégrante. Le processus de « naturalisation de la technique » permet la justification d'un tel projet. Si l'on s'en tient à la thèse de Le Dévédec, le but ultime du transhumanisme est l'affranchissement des limites biologiques dans l'idée d'une émancipation de l'Homme des contraintes que lui impose la nature (Le Dévédec, 2015 :198). En effet, l'idéologie transhumaniste s'établit sur une « dévalorisation complète de l'être humain et de son corps » (Le Dévédec, 2015 :198). Tout ce qui se rapporte au biologique relève d'un handicap à la perfectibilité humaine. La condition de survie de l'humanité dépendrait alors de l'émancipation de l'Homme de ses frontières biologiques aliénantes. Dépasser les limites du possible, contourner les lois de la nature afin de créer un Homme nouveau ; un « posthumain » dont les capacités seront décuplées à l'aide des technosciences, voilà l'objectif du mouvement transhumaniste, en témoigne les propos de Ray Kurzweil :

« As we're learning the principles of operation of the human body and the brain, we will soon be in a position to design vastly superior systems that will be more enjoyable, last longer, and perform better, without susceptibility to breakdown, disease, and aging. »  
(Kurzweil, 2003)

Dans la synthèse que propose Pieter Lemmens de l'ouvrage de Robert Ranisch et Stefan Lorenz Sorgner, *Post- and Transhumanism*, l'auteur explique que le transhumanisme se veut un « ultra-humanisme ». Le courant souhaite donc aller au-delà de l'humanisme sans s'en détacher complètement. Lorsque Ray Kurzweil prédit l'arrivée prochaine de systèmes

supérieurs à celui du corps humain, il ajoute (et on comprend donc que c'est ce qu'il entend par supérieur) : « des systèmes qui seront plus agréables, durables et performants, sans risque de dégradation, de maladies ou de vieillissements. » La dépréciation de l'être humain dans la pensée transhumaniste est très clairement représentée ici. Cette dévalorisation de l'Homme rompt avec le courant humaniste qui voit en l'humain et la société le moyen suprême de l'émancipation. « L'être humain fait exécuter les tâches difficiles par des machines intelligentes. Cette délégation du pouvoir est aussi l'aveu de la supériorité de la machine sur l'intelligence humaine... du moins le croit-on ! » (Andrieu, Anderson, Le Parvis, 1992 : 9). La supériorité de la machine tient au fait que, contrairement à l'Homme, elle n'est pas sensible aux émotions, mais est seulement « intelligente » au sens opérationnel qu'on lui en donne ; soit capable de traiter l'information et de résoudre des problèmes. (Andrieu, Anderson, Le Parvis, 1992 : 8) Dans la pensée transhumaniste, et en IA, la technique est le moyen suprême de l'émancipation. Cela se retranscrit profondément dans la pensée d'Haraway à travers son mythe du cyborg.

« La médecine moderne, elle aussi, fait appel à des cyborgs, accouplements entre organisme et machine, tous conçus comme des systèmes codés, et dont l'intimité et l'énergie ne proviennent pas de l'évolution de la sexualité telle que nous la connaissons. Le sexe cyborgien fait revivre quelque chose de la ravissante liberté répliquative des fougères et des invertébrés (quelle délicieuse prophylaxie naturelle contre l'hétérosexisme.) » (Haraway, Allard, Gardey et Magnan, 2007 : 30)

Elle voit en la technique (plus spécifiquement en la figure du cyborg) une possibilité de libération ; une libération du corps sexué biologique restreint, brimé et asservi par le capitalisme patriarcal.

Ray Kurzweil, par le biais de ses discours, démontre bien l'ambition transhumaniste d'une hybridation de l'être humain avec la technique, dans laquelle cette dernière dépasse largement le biologique et permet par ce dépassement de surélever l'Homme afin de parvenir à une « meilleure » version de lui-même : « Human Body Version 2.0 », comme l'indique le titre du billet de blogue (Kurzweil, 2003).

« By 2040, the nonbiological portion of our intelligence will be far more powerful than the biological portion. It will, however, still be part of the human-machine civilization, having been derived from human intelligence, i.e., created by humans (or machines created by humans) and based at least in part on the reverse-engineering of the human nervous system. » (Kurzweil, 2003)

On remarque selon la conception de Kurzweil, que l'humain dans sa forme biologique ne serait qu'une « version » première qui suppose une version élevée à venir au même rang que n'importe

quel objet technologique moderne. Plus encore, ce discours de Ray Kurzweil démontre qu'il ne fait aucun doute selon les transhumanistes que l'IA va un jour surpasser l'intelligence humaine. Afin de comprendre ce que cela implique d'avancer de telles promesses, il est essentiel de comprendre la conception de l'intelligence des transhumanistes. C'est là tout l'intérêt de la démarche de notre projet de recherche. Comme on l'a vu, l'intelligence est un concept construit et leur définition de l'intelligence est réduite à une conception logico-mathématique. Dans cette perspective, le dépassement de l'intelligence humaine par l'IA signifierait seulement une meilleure performance de la machine dans sa capacité de traitement de l'information afin d'optimiser sa performance en vue d'un objectif.

Dans une perspective constructiviste, nous avons vu que l'intelligence se comprend comme un concept social construit selon une représentation de la réalité propre à un univers épistémologique particulier : le modèle informationnel. À l'intérieur même de ce modèle, plusieurs définitions de l'intelligence vont germer au fil des stades de développement de l'IA au regard de la finalité attendue des programmes. De la rétroaction à l'optimisation en passant par la résolution de problème, chacune des finalités escomptées entend une conception différente de l'intelligence par un traitement de l'information. À l'image de l'Homme réduit à son intelligence dans sa fonction opérationnelle, l'IA représente bel et bien cette créature artificielle tant fantasmée. La distinction de la nature et de la technique ne signifiant désormais plus rien lorsque le fonctionnement de toutes entités confondues s'explique par le même processus informationnel, le domaine de recherche technoscientifique est ainsi légitimé.

La stratégie de traduction (du langage naturel au langage formel), la simplification des méthodes (prise en compte seulement du mécanisme global et non complexe de la biologie humaine), et la manœuvre d'élimination (de certains aspects intégrants de l'être humain selon d'autres représentations de l'intelligence) élaborée dans l'optique de simuler l'intelligence dans l'ordinateur, démontrent le réductionnisme des chercheurs en IA. Cette perspective restreinte légitime le projet de l'IA et justifie la possibilité d'une telle démarche. Au vu de la lignée ancestrale des créatures artificielles façonnées à l'image de l'Homme dans laquelle s'inscrit l'IA, la volonté de reproduction de l'être humain n'est pas illégitime en soi. La réduction de l'être humain à son intelligence et de l'intelligence au simple traitement de l'information en IA à des fins opérationnelles pose, en revanche, un problème ontologique majeur. Autant dans l'idéologie que dans le concret, ce projet est représentatif de la dépréciation de la nature humaine ; du biologique au détriment de la technique. En opposition avec la pensée humaniste,

l'IA rend compte d'une relève des moyens techniques pour sortir l'humain de l'aliénation de sa propre condition biologique. L'Intelligence Artificielle représente, aussi bien dans ces (ses termes que ses applications, le reflet d'une société qui met en avant la rationalité technique au-devant de la pensée humaine et collective. La confiance accordée à la logique mathématique, aux chiffres, aux algorithmes pour prendre le contrôle de notre quotidien au-delà des émotions humaines, de l'instinct, de la signification de nos paroles et de l'intentionnalité de nos gestes dévoile un portrait révélateur de notre société technicisée. Créée selon notre représentation de l'être humain, l'IA exerce à son tour une influence notable sur notre façon de se représenter l'Homme. L'IA a le pouvoir de nous faire reconsidérer les frontières de l'être humain et en donne ainsi une nouvelle définition.

## **Chapitre III.**

### **La démarche méthodologique**

Nous avons établi lors des deux précédents chapitres une réponse théorique à notre problématique de base. C'est empiriquement que nous étudierons à présent la question. Ce présent chapitre consiste donc à détailler de manière concise et précise les différentes étapes de la démarche méthodologique.

Nous croyons que l'analyse de discours est la méthode qui manifeste le mieux les représentations sociales de l'intelligence. Phénomène social important en cette décennie, l'IA est au centre des débats médiatiques. Les discours concernant la discipline scientifique de recherche se multiplient nous procurant un corpus vaste et complet sur lequel appuyer notre étude qualitative. Des entrevues avec les grands chercheurs en IA jusqu'aux pages web des GAFAs mettant en valeur les bienfaits de la recherche et du développement de l'IA au sein de leur entreprise ; les matériaux mis à notre disposition offrent une fenêtre privilégiée pour investiguer la manière dont est représentée l'intelligence. Nous trierons et sélectionnerons ces matériaux afin d'avoir un corpus représentatif de diverses formes discursives qui, à la fois diffèrent par leur provenance et angle de vue, et à la fois partagent une même constante : un discours de vulgarisation traitant de l'IA.

À travers cette analyse, nous chercherons à dévoiler les représentations sociales dominantes retranscrites dans les discours. Nous ne nous intéresserons qu'aux discours de vulgarisation. Dans la traduction du langage scientifique au langage profane paraît le plus clairement la définition de l'intelligence de l'émetteur du discours.

Après avoir présenté le phénomène social qu'est l'IA et avoir défini notre objet de recherche, il s'agira d'expliquer en quoi consiste la méthode qualitative de recherche que constitue l'analyse de discours pour enfin s'attarder sur la construction des données sociologiques.

#### **1. L'objet de recherche**

##### **1.1. L'Intelligence Artificielle en tant que phénomène social.**

Il est possible de considérer l'IA comme un phénomène social à partir du moment où, celle-ci, pas seulement en tant que domaine de recherche scientifique, mais en tant que

technoscience, a un réel impact sur la société et bouleverse ainsi notre rapport au monde. Bien que l'usage de l'IA soit plus courant que ce que nous imaginons, ce n'est pas seulement l'impact de cette technique sur notre quotidien qui créer ce bouleversement, mais la place de plus en plus importante qu'elle occupe dans l'imaginaire collectif. Cette place résulte de l'engouement récent envers l'IA retranscrit à travers les médias, l'intérêt des gouvernements (via les stratégies politiques mises en place pour et autour de l'IA), les investissements massifs publics (via les importantes subventions allouées à la recherche universitaire) et privés (par l'entremise des GAFAs, autres multinationales, et multiples start-up).

L'IA a pour cible tous les secteurs de l'activité économique lui octroyant ainsi le titre de « véritable révolution industrielle » selon les discours de nombreux acteurs sociaux. L'IA entend en effet bouleverser les secteurs du travail, de la santé, de la sécurité, de l'écologie, de l'humanitaire, de l'urbanisme et tout ce qui occupe et guide notre quotidien. L'IA constitue une promesse dans le sens où elle se veut être la réponse aux maux du monde. C'est d'ailleurs une des raisons pourquoi elle suscite de nos jours un tel engouement. C'est au nom du *progrès*, en tant qu'amélioration de la condition humaine par le biais de la technique, qu'est légitimé le développement de l'IA, comme bien des technosciences avec elle. L'IA est vendue au travers des discours de vulgarisation comme un produit répondant à toutes les questions et à tous les problèmes de notre planète. Les technosciences, et d'autant plus l'IA caractérisée par sa possibilité d'application dans presque tous les domaines, permettent de se débarrasser de la responsabilité des ravages de notre monde. Il semble plus facile de laisser le sort de notre existence à la technique plutôt que d'en endosser la responsabilité, ce qui impliquerait faire face à nos échecs, remettre en question nos a priori et, ultimement, devoir renoncer à nos habitudes et à notre confort. Dans un sens, donc, le phénomène social que constitue l'IA constituerait le remède le plus efficace et le plus rapide contre la déperdition de notre condition humaine sur le court terme tout en participant sur le long terme à cette dépravation. En effet, n'empêche-t-elle pas d'une certaine manière le développement d'une réflexion de fond sur notre mode de vie, essentielle pourtant au changement de perspectives envers notre monde et à la reconfiguration de celui-ci ? De toute évidence, elle l'a ralenti.

Nous avons exploré au cours de ces deux premiers chapitres l'aspect théorique de notre problématique : la représentation sociale de l'intelligence en IA fondée sur le modèle informationnel. L'objectif du prochain chapitre est d'étudier si cette représentation informationnelle de l'intelligence se retranscrit à travers les discours de vulgarisation en IA et si oui par quels moyens. C'est donc par une analyse discursive que nous chercherons à évacuer

la représentation centrale de l'intelligence à la base de notre rapport au monde et du développement autant idéologique, que scientifique de l'IA.

Par l'entremise de discours provenant de sphères diverses, nous tenterons de dévoiler les représentations sociales de l'intelligence en IA. Cela permettra de confirmer ou de réfuter empiriquement notre postulat établi théoriquement selon lequel ces définitions se rassemblent sur le fait qu'elles reposent toutes sur une conception informationnelle du monde et sont irréductiblement réduites à un simple traitement de l'information permettant l'optimisation.

Dans l'exercice de vulgarisation scientifique établi via une traduction de leurs travaux informatiques et mathématiques en langage profane, les chercheurs en IA transmettent clairement leur vision de l'être humain et de son intelligence. Les discours publicitaires, et en particulier celui des GAFAs, cherchent à vendre cette représentation par les biais de la communication et du marketing. Les discours politiques, quant à eux, confèrent une certaine autorité à cette vision de l'être humain en élaborant autour du domaine une stratégie politique bien définie. Le discours transhumaniste, enfin, est celui qui soulève le plus explicitement la représentation de l'intelligence dans le modèle informationnel ; leur but étant de divulguer l'idéologie transhumaniste à travers l'exercice intellectuel. Le discours journalistique entretient cette représentation informationnelle de l'intelligence tout en la divulguant via les différents médias, ou la critique. C'est pourquoi nous avons choisi de ne pas nous limiter à un certain type de discours tant la diversité de provenances et d'acteurs sociaux signataires de ceux-ci permet de rendre compte d'une vision globale commune de l'être humain et de son intelligence.

## **1.2. La définition de l'objet de recherche : l'intelligence en IA.**

L'analyse de discours (chapitre 4) aura pour objectif de soulever la ou les représentation(s) de l'intelligence que l'on retrouve(nt) dans les différents types de discours touchant l'IA. Ces parenthèses alternatives ne sont pas laissées au hasard. En effet, l'analyse discursive nous indiquera si le postulat qui est le nôtre est confirmé ou réfuté. S'il est confirmé, cela signifie qu'il existe une seule et même représentation sociale de l'intelligence en IA montrant que cette conception informationnelle de l'intelligence dépasse le seul entendement de l'auteur du discours, de son champ disciplinaire respectif ou même de son idéologie. Cela démontrerait que cette conception est justement prise dans une configuration plus large et englobante du monde formulée à partir du paradigme informationnel. En revanche, s'il est réfuté, cela montrera que cette représentation soulevée théoriquement ne se retranscrit pas dans

tous les discours de vulgarisation sur l'IA et que donc, l'influence du paradigme informationnel sur les conceptions symboliques n'est pas si importante et globalisante qu'elle n'y paraît. Il serait donc relativement facile de s'extraire de ce modèle.

Quelle que soit la conclusion de notre recherche, les représentations sociales de l'intelligence se trouvant dans les discours ambiants constituent notre objet de recherche. Il s'agira d'établir en quoi ces représentations mettent en avant une certaine compréhension du monde ; une compréhension, s'il l'on s'en tient à notre analyse théorique, extrêmement réductrice et opérationnelle. Nous verrons en quoi les conceptions symboliques véhiculées à travers les discours ambiants participent au renforcement d'une vision unique qui légitime le développement de l'IA. Les discours de vulgarisation de l'IA tendent à valoriser une certaine représentation spécifique de l'intelligence et, par-là, à transmettre une certaine image de l'Homme qui participe à la reconfiguration des frontières de celui-ci.

Que ce soit Yoshua Bengio (chercheur), Google (entreprise), Cédric Villani (politique), Ray Kurzweil (transhumaniste) ou *La Presse* (journal), tous tentent d'élaborer des réponses à ces questions (particulièrement dans le discours de vulgarisation) : qu'est-ce que l'IA, à quoi sert-elle et quels enjeux lui sont rattachés ? En effet, il est question de donner un sens à cette technoscience qui prétend reproduire artificiellement l'intelligence. Cette dernière étant, selon nous, un concept construit socialement ; il s'agit alors de mettre en mot leur représentation respective de l'intelligence. La vocation de notre analyse sera de soulever la définition donnée à l'intelligence au sein de chacun des discours analysés ; puis de divulguer la ou les représentation(s) sociale(s) qui en découlent.

Nous comprenons le concept de *représentation sociale* selon la définition de Denise Jodelet : « connaissance, socialement élaborée et partagée, ayant une visée pratique et concourant à la construction d'une réalité à un ensemble social » (Jodelet, 2003 : 53). C'est selon ce même angle d'approche que ce se basera notre analyse de discours.

## **2. L'analyse de discours**

L'analyse de contenu constituera la méthode mobilisée dans la partie empirique de ce présent mémoire. Nous étudierons le contenu oral et écrit de discours dont le propos touche à l'IA, tout en prenant compte dans l'analyse le contexte dans lequel chacun de ces discours s'inscrit.

## 2.1. Méthode de l'analyse discursive

Alors que les données (« data ») qui nourrissent l'IA peuvent exister en elles-mêmes physiquement comme entité singulière, les données en sciences sociales que nous étudions sont des données « signifiantes » qui se doivent d'être éclairées à la lumière d'une analyse sémantique de contenu puisqu'irréremédiablement rattachées à leur contexte spatio-temporel.

[...] les sciences de l'homme se distinguent des sciences de la nature en ce qu'elles font intervenir un type de données irréductible aux données du monde physique : ces données sont signifiantes. Comme le dit R. Aron (1948 : 57) à propos de l'histoire : « La science du passé humain bénéficie d'un privilège qui lui appartient à elle seule, elle s'attache à des êtres qui ont pensé et dont elle veut reprendre la vie ou la conduite. » (Molino, 1985 : 281)

Dans la partie empirique de ce mémoire, nous cherchons à dévoiler le sens implicite des propos retrouvés dans les discours centralisés sur le thème de l'IA. Ces significations, soulevées à travers l'analyse de discours, permettent de « mettre en relief la nature [des] relations, en plus de confirmer l'hypothèse de l'existence de ces relations. » (Negura, 2006 : 9)

Selon la définition de Laurence Bardin, l'analyse de discours constitue un « ensemble de techniques d'analyse des communications visant, par des procédures systématiques et objectives de description du contenu énoncé, à obtenir des indicateurs permettant l'inférence de connaissances relatives aux conditions de production/ réception de ces énoncés » (Bardin, 2013 : 43). Nous procéderons donc à une analyse de contenu explorant la dimension discursive des conduites sociales par le moyen d'une démarche méthodologique spécifique. Différentes approches peuvent être empruntées en analyse de discours. Nous choisirons pour ce présent mémoire d'étudier notre objet de recherche via l'angle « analytique ». Cette approche vise à étudier les différentes formes de connaissances constitutives d'un phénomène social identifié. Il n'est donc ni question de restituer (comparer les représentations sociales du discours avec la conception normative de la vie sociale), ni d'illustrer (sélectionner sans justifier) les propos tenus dans les discours. Nous emprunterons une perspective « linguistique » ou encore appelée « sémantique » au fil de cette analyse. L'usage de cette approche comprend le langage comme base d'appréhension du matériau discursif afin de pouvoir en soutirer une analyse objective fondée sur les formes symboliques de la sémantique.

Au même titre que tout agent actant, les discours structurent la réalité sociale ; c'est pourquoi le matériau discursif est essentiel à inclure dans l'analyse sociologique, car il ne se

contente pas de décrire et d'illustrer un phénomène social, mais agit sur celui-ci. Autrement dit, le discours en lui-même agit sur le social, influençant ainsi les comportements et les décisions des acteurs. Le sens ne se trouve donc pas toujours dans le texte, mais dans la relation entre le texte, le récepteur et le producteur du discours dont le rôle est de mettre en lumière la distinction de sens. L'analyse de discours dans une approche sémantique est une forme d'analyse de contenu qui s'intéresse à la portée symbolique et linguistique d'une production langagière, soit de discours oraux et écrits. Quel que soit le document sélectionné (texte, vidéo, etc.) selon des critères préétablis lors de notre analyse (exposer lors de la construction des données), c'est le contenu discursif, et seulement celui-ci, qui sera la cible de l'analyse. Les discours dévoilent le sens et la symbolique d'une production indissociable de son auteur et de sa visée. Ils sont structurés par la réalité sociale tout en produisant à leur tour des formes symboliques structurant la vie sociale.

## **2.2. La vulgarisation scientifique**

La sélection des matériaux analysés ne sera pas laissée au hasard. Nous nous concentrerons sur les discours de vulgarisation scientifique. « Le vocable « vulgarisation » est couramment usité en langue française pour désigner toute activité de communication de la science en direction du grand public. » (Bensaude-Vincent, 2010 : 2) Si la représentation de l'intelligence est présente dans les articles scientifiques, elle s'étend et se généralise à travers la vulgarisation. C'est elle qui contribue à la divulgation de la conception de l'intelligence. De plus, elle y est plus apparente à travers cette forme de discours. En effet, dissimulée sous l'épais voile du jargon scientifique, la représentation de l'intelligence se dévoile plus aisément à travers l'exercice de traduction du discours expert destiné au profane que dans le discours expert en lui-même.

Le sens d'un document se trouve autant dans la sémantique (le signifié, le fond) que dans la syntaxe (signifiant, la forme). La syntaxe implique les outils mobilisés pour s'exprimer à l'oral comme à l'écrit, l'appropriation des éléments de la langue et l'usage de ceux-ci selon le contexte spatio-temporel, la visée (à qui s'adresse le discours, mais aussi dans quel but), le lieu de divulgation du discours, l'expérience personnelle et le milieu social de l'auteur. La vulgarisation fait partie de ces stratégies d'appropriation d'une forme de langage spécifique. Celle-ci demande un réel effort de la part de celui qui l'exerce. Des outils de langage très concrets sont d'ailleurs mobilisés : l'anecdote, l'exemple, l'humour, la comparaison, l'analogie,

la métaphore, la formule, la reformulation et la fonction métalinguistique (définition accessible d'un concept technique).

Le discours vulgarisé s'adresse à un public profane dans le but de démocratiser la recherche scientifique pour faire de la production scientifique un bien commun. Elle est divulguée à travers les espaces médiatiques (télévision, presse, revues), la publicité, les rapports gouvernementaux ou lors de conférences. Sous le couvert d'une visée démocratique axée sur la collaboration entre la communauté scientifique et citoyenne dont la recherche d'une élévation culturelle de la population est le seul intérêt, c'est très souvent des raisons marketing qui motivent véritablement l'effort de vulgarisation.

C'est sous une lumière inhabituelle que Bernadette Bensaude-Vincent analyse le concept de vulgarisation dans l'article « Splendeur et décadence de la vulgarisation scientifique » (Bensaude-Vincent, 2010). À travers un retour sociohistorique de la notion, elle critique les fondements et dévoile les effets pervers dissimulés par une visée qui se dit « neutre ». Elle explique qu'historiquement, la vulgarisation n'aurait pu avoir lieu sans l'établissement volontaire d'un fossé entre l'élite scientifique et « la masse d'ignorants » (Bensaude-Vincent, 2010 : 4) . La science, alors isolée et sacralisée, utiliserait la vulgarisation pour « informer » tout en maintenant les citoyens comme éléments passifs ne pouvant influencer le cours de la science et de la technique. « La vulgarisation ne s'adresse pas qu'à l'entendement et ne vise pas nécessairement à aiguïser le jugement » (Bensaude-Vincent, 2010 : 5), mais « elle [...] invite le public à « consommer » la science sous forme de magazines, livres ou spectacles plutôt qu'à s'exercer à la pratiquer » (Bensaude-Vincent, 2010 : 5). La vulgarisation configure, d'après l'auteur, non seulement le public, car elle permet de « faciliter l'acceptabilité sociale des innovations techniques » (Bensaude-Vincent, 2010 : 7), mais aussi la science en elle-même.

« Le style propagande pour la science est d'autant plus vivace que la recherche fait de plus en plus appel à des commanditaires pour son financement. Assumée parfois par les chercheurs eux-mêmes, la communication scientifique entre dans une économie de la promesse : elle laisse entendre que les laboratoires sont sur le point de résoudre tous les problèmes de santé ou de société dans un bref avenir. Mais, si la logique de participation du public entre un jour pleinement en jeu, elle devrait conduire à privilégier la diversité des options scientifiques et technologiques plutôt que l'investissement dans une trajectoire unique, jugée optimale et toujours présentées comme « la révolution » qui va résoudre une foultitude de problèmes. » (Bensaude-Vincent, 2010 : 9)

### **2.3. L'aide à l'analyse qualitative via le logiciel Atlas.ti**

Nous effectuerons l'analyse de discours à l'aide du logiciel d'analyse qualitative Atlas.ti. Avant de procéder à notre recherche, il est essentiel de produire une pensée réflexive sur sa démarche méthodologique et de prendre en considération les portées d'un tel logiciel, ces bénéfiques, mais aussi ses limites afin que la partie empirique de ce mémoire soit scientifiquement valide.

Atlas.ti est un logiciel informatique ayant pour but d'assister le chercheur dans sa recherche en l'aidant dans l'analyse de contenu qualitatif (textes, vidéos, images, etc.). Le programme fournit de multiples outils permettant de coder, localiser et annoter les données qui constituent la base empirique à l'interprétation théorique. Il permet de mettre en perspective les composants et significations d'un discours, d'une image ou même d'un contenu vidéo afin d'évaluer, comparer et peser leur importance. Interpréter les significations cachées d'un contenu qualitatif plus rigoureusement ; tel est l'objectif principal d'un programme de Méthode de la Théorisation Enracinée (MTE).

Le logiciel n'est pas conçu dans l'optique de remplacer le travail du chercheur dans l'analyse et l'interprétation du contenu ; mais fait office d'outil d'assistance. « Le logiciel n'est qu'une interface permettant de faciliter ce processus qui était autrefois réalisé de façon manuelle. » (Bandeira de Mello et Garreau, 2011 : 185) Alors que le chercheur devait auparavant coder, classer et annoter lui-même à la main les multiples documents à analyser ; il peut, grâce au programme, centraliser l'ensemble de ses réflexions sur un même fichier, réduisant ainsi considérablement le temps à classer et les risques de pertes de pistes d'interprétations possibles (Bandeira de Mello et Garreau, 2011 : 183). La fonction « recherche » permet de récupérer automatiquement le contenu qui nous intéresse sans rien ne laisser au hasard. L'analyse gagne ainsi en validité scientifique (Bandeira de Mello et Garreau, 2011 : 189). Le codage ouvert nous permet de catégoriser les éléments des discours sélectionnés par nos soins avec flexibilité et clarté afin de pouvoir souligner les composants soulevant la représentation de l'intelligence dans les discours des acteurs de l'IA (Bandeira de Mello et Garreau, 2011 : 185).

Rodrigo Bandeira de Mello et Lionel Garreau dans l'article « L'utilisation d'Atlas.ti pour améliorer les recherches dans le cadre de la Méthode de la Théorisation Enracinée (MTE) : panacée ou mirage ? » mettent en gardent contre les pièges auxquels peuvent être confrontés les chercheurs dans leur utilisation du logiciel d'aide à l'analyse qualitative des données ; notamment celui de réifier les codes (se contenter de manipuler des codes pour faire avancer

l'analyse en leur accordant un statut objectif et donc indiscutable), rendant ainsi la recherche « superficielle » (Bandeira de Mello et Garreau, 2011 : 191). De plus, le logiciel ne serait méthodologiquement neutre puisque construit selon la logique de ses créateurs. Il s'agit donc de se plier aux fonctions, ressources et méthodes que ces chercheurs pensent optimales (Bandeira de Mello et Garreau, 2011 : 193). En s'adaptant à cette méthode imposée par le programme, le jeune chercheur n'est pas poussé à développer une réflexion sur ses choix méthodologiques (Bandeira de Mello et Garreau, 2011 : 194).

Une des stratégies proposées par les auteurs pour éviter les potentiels pièges est d'utiliser seulement les fonctions du logiciel qu'il pense bénéfique à sa recherche. Ainsi, nous utiliserons pour ce présent mémoire uniquement la catégorisation et le codage. C'est donc en maximisant les bénéfices tout en gardant en tête les risques potentiels que nous nous épaulerons d'Atlas.ti dans la construction de notre analyse de discours.

### **3. Construction des données sociologiques**

#### **3.1. Identification du matériau d'analyse**

L'analyse de discours à laquelle nous aurons recours permet d'entrevoir l'étendue de la définition de l'intelligence en IA. Celle-ci n'est pas seulement représentée dans les dires des chercheurs en IA. C'est pourquoi nous avons étendu notre sélection à d'autres protagonistes dont le discours traite de l'IA. Analyser à la fois les discours des acteurs directs du domaine technoscientifique (chercheurs, transhumanistes) et ceux des acteurs externes indirects (journalistes, gouvernementaux) permet de rendre compte du monisme réductionnel de cette définition. Celle-ci est partagée par les différents protagonistes ; alimentée et soutenue par certains (chercheurs, transhumanistes et entreprises) et divulguée par d'autres (journalistes, gouvernements). L'intérêt de cet ajout pour notre recherche est de démontrer que les chercheurs ne sont pas seuls à la base de cette acquisition de définition de l'intelligence. L'autorité que leur apporte leur statut influence considérablement les sphères en dehors du secteur scientifique, et la divulgation de cette définition par ces derniers permet la légitimation de la représentation sociale assignée au concept. Nous verrons aussi que les différents types de discours s'entrecoupent. Les assigner à une catégorie spécifique ne relève donc pas toujours de la facilité. Nous tenterons néanmoins de nous atteler à cet exercice dans un but de clarté et nous nous appliquerons à souligner ce chevauchement.

Seuls les discours émis ou publiés après les années 2000 seront l'objet de notre analyse. Ce tri volontaire permet de focaliser notre analyse sur des discours contemporains. C'est à partir des années 2000 que l'on a tenté de faire ressurgir l'IA de son hibernation. La discipline ou technoscience qui en découle est alors devenue sujet d'actualité. La démocratisation de l'IA à travers l'avalanche de discours qui l'entourent a entraîné des investissements plus conséquents et avec eux une reprise fortuite de la recherche. La prolifération de matériel pour notre analyse est alors très dense, malgré cette limitation aux discours contemporains. De plus, cette sélection permet de diriger l'attention de notre analyse sur la définition contemporaine de l'intelligence en IA dont la dimension de l'optimisation est centrale.

### *Le discours des chercheurs*

Une grande partie des discours analysés (18 sur 39) proviennent des chercheurs en IA ou toutes autres disciplines connexes relativement proche et en rapport : les scientifiques et analystes des données, informaticiens, architectes de logiciel et chercheurs en neurosciences. L'IA est cependant au cœur du sujet de chacun des discours analysés, quelle que soit la discipline d'origine de son émetteur. Certains sont des chercheurs universitaires affiliés au domaine public : Yoshua Bengio rattaché à l'Université de Montréal, Marvin Minsky, Alex Wissner et Newton Howard au MIT et Jean-Gabriel Ganascia, chercheur en IA à la Sorbonne et Paris 6. Ce dernier est surtout connu pour ses réflexions philosophiques concernant l'IA. Tout comme ce dernier, Gilles Dowek (autre chercheur affilié à la sphère universitaire) est à la fois informaticien et philosophe. Toutefois, la plupart sont rattachés à la sphère privée et sont chercheurs pour des laboratoires appartenant aux entreprises : Joëlle Pineau, directrice du laboratoire de recherche de Facebook à Montréal ; Yann LeCun, scientifique en chef pour Facebook AI Research (FAIR) ; Margaret Mitchell, chercheuse en IA au laboratoire GoogleAI. D'autres appartenant aux disciplines connexes travaillent pour des laboratoires externes, mais aussi privés : Tom Gruber fondateur de Siri ; Poppy Crum, chercheuse en neurosciences spécialisée en IA ; Mallory Freeman et Jeremy Howard en tant que scientifiques des données, Joseph Redmon, Blaise Aguerre y Arcas et Grady Booch, architectes de logiciel d'IA et Kevin Slavin, expert en algorithmes d'IA. Nous avons aussi choisi d'intégrer à l'analyse un discours de Jacques Testart, même s'il n'est en rien affilié à l'IA tant dans son secteur de recherche (étant biologiste de formation) que dans ses convictions. Il lui arrive toutefois de s'exprimer sur le sujet en se positionnant à l'encontre de la pensée transhumaniste, notamment dans le discours analysé « L'abeille est plus intelligente que l'ordinateur », dont le propos ne pourrait être plus pertinent à notre analyse.

La plupart des chercheurs cherchent à maintenir une certaine distance avec les transhumanistes. Ils n'hésitent d'ailleurs pas à montrer leurs différends par la critique des ambitions extrêmes de ce courant de pensée. Les chercheurs, à travers leurs discours de vulgarisation, souhaitent montrer que l'IA ne constitue rien de plus qu'une technologie utilitaire allant dans le sens du progrès. Il n'y aurait donc aucune raison de craindre que son développement puisse refaçonner les frontières de l'être humain.

Si les discours diffèrent et que la distance est volontairement maintenue, nous verrons que les chercheurs et transhumanistes partagent pourtant dans les faits une même vision de l'Homme et de son intelligence. Nous retrouvons en effet chez nombre de chercheurs cette dépréciation de la nature biologique de l'être humain face à l'élévation du progrès technique, très représentative de la pensée transhumaniste. De ce point de vue, le dépassement de l'Homme par la machine leur semble à eux aussi inévitable. Certains discours de chercheurs (Margaret Mitchell, Tom Gruber) rejoignent très clairement la pensée du courant transhumaniste. Leur foi inébranlable en la technique constitue la marque première de leur assimilation à la pensée transhumaniste. Ne s'identifiant pas au mouvement transhumanisme, nous avons néanmoins pris la décision de les garder dans la catégorie des chercheurs. Pour les raisons évoquées plus haut, il est parfois difficile de distinguer un discours d'un chercheur de celui d'un transhumaniste. La frontière entre les deux se trouve pour nous dans l'affiliation assumée au courant de pensée. L'impact du discours sur son auditoire ou son lecteur ne sera pas le même selon l'affiliation de l'auteur au mouvement ou non. En tant que chercheur et non-membre de l'association, ils se présentent avant tout en tant que scientifiques. L'autorité est alors plus forte et le discours ne sera pas reçu de la même manière et avec la même attention. En effet, les transhumanistes peuvent perdre en légitimité lorsqu'ils évoquent leurs grands projets futuristes technophiles.

### *Le discours transhumaniste*

De l'autre côté, certains auteurs assumant leur affiliation au mouvement transhumaniste, se trouvent être des chercheurs en IA : Max Tegmark, professeur et chercheur au MIT, et Sam Harris, chercheur en neurosciences.

Le but des transhumanistes est de convaincre des possibilités que peuvent offrir l'IA à l'amélioration de la condition humaine et non de chercher des moyens de développer concrètement ces moyens. À la différence du chercheur qui développe des algorithmes en IA dans une logique utilitaire, dont l'optique est de développer des outils au service d'un quotidien moins aliénant pour l'Homme, le transhumaniste entrevoit dans l'IA la voie à la transformation

de la nature de l'être humain. Parce que la visée diverge, le discours sera élaboré différemment. La divulgation de leur idéologie technoscientifique passe inévitablement par la formulation explicite de leurs ambitions. Dans la lignée des premiers chercheurs en IA symbolique, il est essentiel pour les transhumanistes que le développement de l'IA soit porteur d'assises idéologiques. Ainsi, l'établissement théorique de ces fondements n'est point négligeable à la différence des chercheurs contemporains en IA dont l'ambition de développer une théorie de l'esprit humain a été délaissée au détriment d'une recherche d'applications optimales. C'est pourquoi les représentations de l'intelligence des transhumanistes nous semblent pertinentes à analyser en addition de celles du chercheur en IA.

La sélection des sujets transhumanistes a été établie sous les critères de leur appartenance à des groupes et associations ayant pour but l'utilisation de la technologie et de la science à des fins d'amélioration du potentiel humain, soit : Humanity+ (l'association internationale et officielle des transhumanistes) et Future of Life Institute (organisation concentrant ses recherches sur les moyens d'exploiter la technologie à des fins positives pour l'avenir). Ces deux institutions regroupent des figures importantes autant dans le monde scientifique (Ray Kurzweil, pape du transhumanisme, mais aussi ingénieur principal chez Google), dans l'entrepreneuriat (Elon Musk, PDG de SpaceX, Tesla et Neuralink), que dans les sciences humaines (Nick Bostrom, philosophe fondateur de Humanity+). Ils jouissent aussi d'une influence considérable en raison de leur haut statut social dû à leur capital économique et culturel élevé.

### *Le discours d'entreprises*

La valorisation du discours du chercheur se reflète parfaitement dans les discours des entreprises dont le but est de développer des applications à partir d'algorithmes d'IA. Loin de ressembler aux discours publicitaires habituels, mettant simplement à l'affiche leur produit, les entreprises basées sur l'IA mettent davantage en valeur sur leur site Web leur secteur de recherche, leur vision éthique et la visée générale de leur démarche. Pour cela, elles empruntent une vulgarisation scientifique propre aux chercheurs se voulant, à première vue, plus informative que marketing. Il n'est pas rare que l'émetteur du discours ne soit autre que le chercheur affilié à l'entreprise lui-même ; justement en raison de l'autorité reconnue de son statut. Cette stratégie permet un détachement avec tout autre produit mis sur le marché.

Nous avons surtout focalisé notre analyse sur les discours de l'entreprise Google (Google AI) et ses affiliations (DeepMind et X. The Moonshot Factory), principale

multinationale en IA. Nous avons aussi analysé les discours de FAIR (laboratoire de recherche de Facebook) ainsi que MILA (Institut des Algorithmes de Montréal).

### *Le discours gouvernemental*

Les rapports gouvernementaux ont été rédigés à la demande du gouvernement au service du citoyen. L'ambition étant d'établir une stratégie d'action face aux investissements à venir en IA. Là encore, les écrits sont plus informatifs que politiques et s'inspirent très largement des travaux de chercheurs. Trois rapports gouvernementaux ont été sélectionnés pour notre analyse :

- Celui des États-Unis : « Preparing for the Future of Artificial Intelligence », élaboré en octobre 2016 par le comité en technologie du bureau exécutif du président : le National Science and Technology Council (NSTC).
- Celui du Canada : « Artificial Intelligence in Canada : Where do we stand ? » ordonné par le conseil des technologies en information et en communication (ICTC) et publiées en avril 2015.
- Celui de la France : « Mission Villani sur l'Intelligence Artificielle ». Nous avons analysé le livret de vulgarisation intitulé « Qu'est-ce que l'Intelligence Artificielle » qui résume le rapport officiel rendu public lors de la conférence « AI for Humanity » tenue au Collège de France en mars 2018. Cédric Villani est un informaticien spécialisé en IA et désormais député de la République en Marche, parti politique du président de la République actuellement en fonction, Emmanuel Macron.

Nous avons aussi analysé la « Déclaration de l'IA responsable de Montréal ». Cette déclaration, mise en place en 2018, a pour visée d'établir des grands principes éthiques auxquels doivent se confondre les chercheurs, entreprises et gouvernements pour le développement d'une IA responsable, autonome, juste, démocratique, au service du bien-être et dans le confort de la vie privée de tous (grandes valeurs fondamentales constituant les grandes lignes de la déclaration). Ce rapport collectif a été initié par l'Université de Montréal et a fait appel dans son processus de rédaction à la participation citoyenne dans une visée démocratique.

### *Le discours journalistique*

La dernière catégorie de discours à analyser est celle du discours journalistique. Il se trouve à la croisée des discours des chercheurs et de l'imaginaire des citoyens face à la

technologie. Il permet de divulguer par le biais de la vulgarisation les nouvelles applications, les enjeux à venir, et la situation politique du pays concerné.

Beaucoup d'entretiens de chercheurs se font dans les médias. La définition journalistique de l'intelligence se trouve ainsi soit dans les questions que le journaliste pose au principal concerné, soit dans l'introduction. C'est toutefois dans cette dernière que la représentation de l'intelligence du journaliste se fait le plus entendre. Notre analyse du discours journalistique se concentrera donc essentiellement sur cette partie. Deux articles seulement figurent à part entière dans les discours journalistiques, n'ayant pas pour but d'entretenir un chercheur, mais d'informer le lecteur sur l'IA : « Cinq idées reçues sur l'Intelligence Artificielle » de Elisa Braun pour le journal français *Le Figaro* publié en janvier 2018, et « Intelligence Artificielle : les machines sont-elles nos amies ? », article publié dans l'hebdomadaire français *L'Express* en 2014.

En ce qui concerne les autres articles de journaux, la section de l'interviewé a déjà été analysée dans les catégories chercheurs et transhumanistes. Nous avons donc, cette fois-ci, focalisé notre attention sur le discours de l'intervieweur. Les entrevues mentionnées sont les suivantes : « Entretien entre Philippe-Audrey Larrue St-Jacques et Yoshua Bengio, roi de l'intelligence artificielle » publiée en 2017 dans le magazine québécois *Urbania* ; de la même année dans le journal québécois *La Presse* : « Yoshua Bengio : chercheur en Intelligence Artificielle » ; l'entrevue réalisée par l'émission québécoise *Tout Le Monde En Parle* « L'intelligence Artificielle » avec la sommité mondiale Yoshua Bengio, retranscrite sur la plateforme YouTube ; l'entrevue radiophonique retranscrite en format écrit sur le site web de *France Info* en janvier 2019 de Luc Julia : « Nouveau monde. Luc Julia : "L'intelligence artificielle n'existe pas et la voiture autonome n'existera jamais" » ; l'interview sur la radio *France 24* de Joëlle Pineau retranscrite là encore à l'écrit en format web en 2018 : « "Les machines peuvent être créatives", assure Joëlle Pineau, à la tête du laboratoire FAIR de Facebook à Montréal » et enfin l'entrevue du transhumaniste Nick Bostrom par le quotidien de presse économique français *Les Échos* : Nick Bostrom : « Le but ultime de l'intelligence artificielle doit être la disparition du travail ».

### **3.2. Description des documents analysés**

Maintenant que nous avons identifié les matériaux en les catégorisant par type de discours, il convient désormais de décrire leur nature. Avant cela, nous tenons à préciser que nous avons fait le choix de concentrer notre analyse sur le propos du discours et non sur son

format ou sa forme en général : tournure, structure, ton employé en fonction du destinataire et de son émetteur, etc. Dans un souci de concision et de pertinence, cet exercice nous paraît non essentiel à notre étude. De plus, ce choix méthodologique nous a permis d'approfondir davantage l'analyse du propos afin de faire ressortir la définition de l'intelligence en IA. Nous dirons seulement cela : la visée du discours diffère en fonction de son type. Le discours publicitaire à une visée marketing : il est question de faire adhérer son auditoire à la technologie qui se développe en vue de se frayer une place conséquente sur le marché via la consommation de la population. Le discours journalistique a une visée plus instructive et informative. Le but étant d'éduquer la population afin qu'elle soit consciente des différentes situations et des options qui se présentent alors à elle. Le discours des chercheurs en IA a plusieurs visées selon l'émetteur du discours en question : la visée peut aussi bien être démocratique, qu'informative, marketing ou simplement se trouver dans une recherche de validation auprès de son public et des investisseurs. L'intention du discours gouvernemental est, elle, purement politique. À travers les rapports ou déclarations, ils élaborent des réponses concernant les droits des citoyens afin de les protéger des enjeux éthiques et sociaux. La visée du discours transhumaniste, elle, se distingue en tout point. En tant qu'association et courant de pensée, le discours transhumaniste prend la forme d'un discours idéologique. Le but est bel et bien de convertir l'individu en leur culte voué à la technique. Comme toute idéologie, le transhumanisme tente d'apporter des réponses aux grands questionnements du monde afin de rassurer l'Homme. En effet, l'hybridation de l'Homme avec la technique constituerait une sorte de réponse aux problèmes auxquels est confronté, ou plutôt sera confronté, l'Homme à l'avenir. Dans leurs discours, nous verrons qu'ils n'hésitent d'ailleurs pas à dicter la manière de penser et le mode de vie à adopter. Leur vocation étant de mettre en place un nouvel ordre social au sein duquel l'individu est central au détriment de la communauté.

Chacun de ces types de discours prend une valeur différente en fonction de leur autorité respective. Le discours du chercheur, en tant que scientifique, est celui le plus valorisé, bien au-delà du discours publicitaire des entreprises. C'est d'ailleurs pour cela que le discours du chercheur est repris dans les autres catégories. De par son autorité supérieure, elle permet de légitimer le développement et l'application pratique de l'IA.

### *Entretiens*

Nous avons étudié les discours des chercheurs tout d'abord à travers les entretiens qu'ils ont pu donner à divers médias : presse papier, radio ou encore émission de télévision. Par cet exercice d'entretien volontaire, la visée de divulgation est alors évidente. Cette ambition de démocratisation n'est néanmoins point présente chez tous les chercheurs. C'est pourquoi

certaines noms reviennent souvent, notamment celui de l'universitaire Yoshua Bengio. Ce précurseur de l'apprentissage profond se donne pour mission de démocratiser le plus possible l'IA, pour faire valoir ses travaux et les multiples possibilités qu'engage cette technologie, mais aussi dans une considération éthique. Il constitue le chercheur le plus impliqué dans la recherche éthique pour un développement responsable de l'IA. Beaucoup d'autres chercheurs ne désirent toutefois pas prendre part au débat. Se concentrer sur leur recherche leur permet une distanciation suffisante à la poursuite du développement des algorithmes d'IA. En effet, le fait même de questionner les enjeux éthiques pourrait remettre en question la pertinence à poursuivre leurs travaux. De plus, s'adresser aux médias demande un réel effort de communication face auquel tous, en tant que scientifiques de carrière, ne sont pas à l'aise.

Certains discours transhumanistes analysés constituent aussi des entretiens. La visée de divulgation n'est cependant pas la même que celle du chercheur en IA. Les transhumanistes comme Nick Bostrom ne cherchent pas à démocratiser l'IA pour que le public soit conscient des implications de cette nouvelle technologie. Ils se soumettent à l'entretien médiatique dans le but de propager le plus largement possible (c'est-à-dire, en dehors de leur cercle restreint) leur idéologie. Les médias constituent de ce fait le meilleur médium de propagation.

Tandis que certains entretiens analysés sont écrits (presse écrite), d'autres constituent des entretiens radiophoniques ou encore télévisuels, soit des discours oraux. Si notre analyse se base sur leur retranscription écrite, il est important de mentionner que ces discours sont formulés par leur émetteur de manière plus spontanée. La représentation de l'intelligence qui en découle est de ce fait plus authentique.

### *Conférence TED*

Nous avons aussi sélectionné pour notre analyse de nombreuses retranscriptions écrites de conférence TED (Technology, Entertainment and Design). Les conférences TED sont organisées au niveau international par une fondation à but non lucratif dont l'objectif est de divulguer et partager des idées novatrices touchant aux nouvelles technologies, à la science, au monde des affaires ou plus généralement aux grands « problèmes » du monde. Sous la forme de courtes allocutions (moins de 18 minutes), les conférenciers sont invités à présenter leurs réflexions, perspectives, théories ou encore concepts dans une rhétorique éloquente qui se veut persuasive et influente. Ces discours inspirationnels donnés aux conférences TED sur l'IA sont présentés par des chercheurs, scientifiques, hommes et femmes d'affaires et créateurs d'applications.

### *Billet de Blog*

D'autres discours, comme celui de Ray Kurzweil, sont tirés de billet de blogue sur le web sous le format écrit et sont divulgués par l'auteur lui-même, lui laissant la possibilité de s'exprimer sans contraintes éditoriales.

### *Pages Web*

Les discours des entreprises en revanche ont tous été pris des sites web officiels de chacune de ces entreprises, mettant en avant leur laboratoire de recherche respectif et les applications qu'ils développent à la suite des travaux de leurs chercheurs. Comme nous l'avons vu, nous n'avons pas affaire à un discours publicitaire ordinaire. Très peu graphique, leur page web est composée principalement de longs énoncés instructifs et informatifs concernant leur mission.

### *Rapports officiels et Déclaration*

Les discours gouvernementaux ont été soutirés de rapports officiels adressés à la population et rédigés par les ministères chargés des nouvelles technologies. Le but est de divulguer leurs stratégies d'actions face aux enjeux politiques, éthiques et sociaux et de présenter leur plan d'investissement concernant l'implantation de l'IA au sein du pays respectif.

La déclaration de l'IA responsable de Montréal diffère dans son objectif et dans sa forme. Elle est rédigée sous la forme de points centraux sur lesquels appliquer un développement responsable et éthique de l'IA.

### *Articles de journaux, séquences radiophoniques et télévisuelles.*

La presse écrite, radio, et émission de télévision sont les principaux médiums des discours journalistiques. Tout en voulant rester neutres conformément aux valeurs d'objectivité journalistiques, nous verrons qu'une certaine vision du phénomène est retranscrite à travers les propos des journalistes contribuant à l'établissement d'une définition moniste de l'intelligence dans l'imaginaire de son public.

## **3.3. Mise en relation de l'objet de recherche et des matériaux.**

Ayant identifié puis décrit nos matériaux, il est désormais temps d'expliquer comment nous avons configuré notre cadre analytique de manière à mettre en lumière, empiriquement, notre objet de recherche. Nous avons séparé notre matériau et organisé son contenu à l'aide du

logiciel Atlas.ti. Trois grandes catégories ont été créées à cet effet : la définition contemporaine de l'intelligence en IA, le réductionnisme de cette représentation et enfin la localisation de l'IA entre nature et artifice. Par cette catégorisation, il nous a été possible de coder chacun des discours sélectionnés. C'est en reprenant cette même configuration que nous avons construit par la suite les grandes sections de notre analyse de discours.

La première grande catégorie concerne la définition contemporaine de l'intelligence dans l'imaginaire des acteurs direct et indirect de l'IA. La sélection des discours contemporains de l'IA a permis de confirmer notre postulat sur l'évolution de celle-ci à travers la progression de l'IA. C'est pourquoi l'analyse se concentre principalement sur l'optimisation comme visée principale de l'IA de nos jours, extirpant ainsi toute dimension théorique et idéologique de son objet de base. Nous verrons en quoi la définition de l'intelligence est inévitablement altérée par cette nouvelle dimension centrale. Nous avons décortiqué les propos des auteurs en points sous-jacents : les données comme nouveau mode de représentation du monde et la recherche d'application en vue d'une fin de la théorie. Nous analyserons en dernier plan les propos remettant en question la définition informationnelle de l'intelligence.

La seconde catégorie concerne le réductionnisme de la représentation de l'intelligence à travers la dimension de l'optimisation. Cette section comprend tout d'abord l'analyse de l'indifférenciation ontologique dans les discours en IA. Puis, elle se concentre sur les qualités non modélisables intrinsèques à l'humain. Certes peu représentative, ce dernier point dévoile une position critique du modèle informationnel au travers de certains discours en IA.

La troisième et dernière grande catégorie d'analyse exacerbe la position de l'IA entre nature et artifice. Nous avons codé les propos des sujets autour de ces deux points centraux : l'IA comme créature artificielle à l'image de l'Homme et le renversement du dualisme comme moyen d'existence de la discipline. Ce dernier point se découpe en deux catégories : l'artificialisation de la nature, dans lequel nous analyserons la dimension de la foi en la technique, et la naturalisation de la technique qui dévoile une réflexion centrée sur les enjeux.

Maintenant que nous avons établi la démarche méthodologique, il convient à présent d'effectuer l'analyse de discours et d'en retranscrire notre interprétation et résultat au prochain chapitre.

## **Chapitre IV.**

### **La représentation de l'intelligence à travers les discours centrés sur l'IA**

Ce chapitre sera consacré à l'analyse de discours basé sur la catégorisation établie à l'aide du logiciel Atlas-ti. La codification des discours nous a permis de soutirer trois grands axes qui guideront cette analyse. Nous entamerons par la définition contemporaine de l'intelligence en IA dont la dimension principale est l'optimisation. Nous dévoilerons dans une seconde partie le réductionnisme de cette représentation. Enfin, nous verrons en quoi les propos des sujets révèlent l'impossibilité de concevoir l'IA sans un renversement du dualisme nature/artifice. L'analyse de discours permet d'illustrer notre propos théorique. Le postulat établi aux chapitres théoriques de notre mémoire sera soit confirmé soit réfuté selon les résultats de cette partie.

#### **1. Définition de l'intelligence dans les discours contemporains.**

Nous cherchons dans cette section à soulever la représentation de l'intelligence se trouvant au sein des discours contemporains sur l'IA. En expliquant en quoi consiste l'IA, les auteurs des discours tentent – et ce presque systématiquement – de définir la notion d'intelligence. Plusieurs d'entre eux admettent la complexité à formuler une définition. Cette difficulté éprouvée s'explique selon le logicien et philosophe des sciences Gilles Dowek dans une approche constructiviste :

« Les progrès en matière d'IA imposent-ils de redéfinir le concept même d'intelligence ? Oui. Mais il n'y a ici rien d'exceptionnel. Nous répondons rarement à une question philosophique – Les universaux existent-ils ? L'histoire est-elle une science ? – par oui ou par non. Répondre à une telle question revient avant tout à préciser ce que signifient les mots « exister », « sciences », etc. Bien souvent nous prenons conscience que ces mots ont plusieurs significations qui demandent à être démêlées et, selon que nous considérons l'une et l'autre de ces significations, la réponse à la question est positive ou négative. Aujourd'hui nous posons la question : une machine peut-elle être intelligente ? Cette question nous invite, avant tout, à nous interroger sur la signification que nous voulons donner à l'adjectif « intelligent ». En particulier, nous définissons habituellement d'emblée l'intelligence comme une faculté humaine. Dans ce cas, la réponse est négative par définition. Nous devons donc, pour lui donner un sens, imaginer une autre définition, qui ne soit pas a priori celle de l'intelligence humaine. » (Dowek, 2018 : 247-48)

Gilles Dowek adopte la perspective constructiviste qui est la nôtre pour justifier l'impossibilité d'obtenir une définition univoque de notre objet de recherche. Autrement dit, aucune définition ne peut être appréhendée indépendamment de son contexte social et de la compréhension de l'univers dans lequel elle s'implante. C'est cette définition, à « imaginer » en accord avec les catégories d'analyses qui structurent aujourd'hui notre rapport au monde, qu'il est question de soulever à travers la section suivante.

Au sein des discours des chercheurs en IA, cette complexité à assigner une définition au concept est envisagée dans une tout autre approche. L'intelligence serait compliquée à définir, car il ne serait tout simplement pas *encore* possible d'émettre une définition définitive de l'intelligence puisque l'étude scientifique du fonctionnement humain n'aurait pas *encore* abouti :

« Non, c'est important que les gens réalisent qu'on a encore beaucoup de recherche fondamentale à faire. Il faut investir dans la recherche fondamentale et on sait pas combien de temps ça va prendre avant d'arriver au niveau d'intelligence des humains. Peut-être 10, 50, 100 ans, on le sait pas. Puis la chose importante, c'est que malgré que les machines sont encore très stupides, elles peuvent faire des choses extraordinaires déjà. » (Bengio, 2017c)

Si la définition de l'intelligence ne fait pas consensus, les chercheurs considèrent qu'il est en revanche possible de résumer cette instance en un simple traitement de l'information ; soit en expliquant son processus global dont la compréhension suffit à la reproduction artificielle :

« Mais à partir du moment où nous admettons que le traitement de l'information est la source de l'intelligence, qu'un système de calcul est la base de l'intelligence, et que nous admettons que nous allons continuer d'améliorer ce type de système et que nous admettons que l'horizon de l'intelligence dépasse totalement ce que nous savons aujourd'hui, nous devons admettre que nous sommes engagés dans la conception d'un pseudo-dieu. » (Harris, 2016)

Sam Harris, auteur spécialiste des neurosciences, dévoile ici l'attribution tribulaire de la notion d'intelligence comme simple processus de traitement de l'information. Le terme « admettre » suppose l'existence d'une vérité fondée qu'il n'est pas nécessaire de questionner. Aucune explication sur les assises de cette assignation n'est émise et aucune ne semble d'ailleurs nécessaire tant le modèle informationnel est ancré dans les représentations sociales et semble donc, pour cette raison, incontestable.

### **1.1. De la rétroaction à l'optimisation**

Les définitions de l'intelligence sont multiples et évoluent parallèlement au développement progressif de l'IA. Selon l'acteur, le terme, la visée et l'interlocuteur, la définition évolue et diverge tout en restant calibrée au sein d'un même univers épistémologique.

Nous avons fait le choix de nous concentrer sur les discours contemporains, qui permettent à eux seuls d'illustrer cette évolution. Le détachement entre la conception d'aujourd'hui et celles des pionniers de l'IA est de cette manière plus perceptible.

Pour Wiener, la définition de l'intelligence n'est autre que la définition de la rétroaction. Les premiers systèmes d'IA (Perceptron, etc.) ont été développés dans cette même représentation de l'être humain et de son intelligence. Or, cette définition de l'intelligence intrinsèquement liée à la rétroaction est loin d'être représentative de la vision actuelle des acteurs de l'IA. Les systèmes évoluent en corrélation avec la représentation de l'intelligence et réciproquement. Lorsqu'on s'épanche vers les rapports sur l'IA émis par les gouvernements en vue d'une stratégie d'action, on voit que la rétroaction comme définition est encore utilisée, mais clairement dévaluée.

« Some define AI loosely as a computerized system that exhibits behavior that is commonly thought of as requiring intelligence. » (NSTC, 2016 : 16)

Il est davantage question ici d'expliquer en quoi consiste l'IA. Le système serait selon cette définition un système artificiellement intelligent dès lors qu'il montre des comportements que l'on assigne habituellement à des comportements demandant de l'intelligence. La définition se base sur cette idée de comportement profondément lié à la rétroaction : « systems that act like humans (e.g., pass the Turing test via natural language processing ; knowledge representation, automated reasoning, and learning) » AI-US report. Autrement dit, qui se comporte (réaction observable) comme l'humain. Ils utilisent cependant le terme très équivoque « loosely ». Ce que le rapport des États-Unis dit ici, c'est que si on veut assigner une définition approximative, grossière et donc imparfaite de l'intelligence, on peut lui assigner la définition de la rétroaction.

Afin d'établir des objectifs définis et clairs sur lesquels baser les investissements publics consentis à l'IA, plusieurs comités se regroupent pour discuter de l'impact sur l'économie du pays de cette technoscience censée révolutionner tous les secteurs de l'économie. Une évaluation des enjeux est de rigueur officiellement pour maximiser les bénéfices et minimiser les risques, officieusement afin de rassurer la population sur leur gestion des enjeux à venir des stratégies d'actions à mettre en place. Ouvert à tous, ce type de rapport se doit d'expliquer en quoi consiste l'IA et de poser les bases de son fonctionnement et de ses objectifs. C'est dans l'analyse de cette section que le paradoxe quant à l'évolution de la conception de l'intelligence est apparent. Lorsqu'il est question d'expliquer en quoi consiste l'IA, que ce soit dans le rapport émit par le Canada comme dans celui des États-Unis, la définition de l'intelligence comme capacité de résolution de problème est entretenue.

« Artificial Intelligence (AI) enables machines or the in-built software to behave like human beings which allows these devices to perceive, analyze data, reason, talk, make decisions and act. » (ICTC, 2015 : 2)

Les auteurs de ces rapports, tenus de mobiliser des connaissances sur le sujet au-delà de la question économique, éthique et sociale ont dû, eux aussi, effectuer une revue de littérature. Or, il suffit de se confronter à la tâche pour s'apercevoir que les réponses à ces questionnements de base théorique se trouvent davantage dans la littérature des pionniers de l'IA symbolique que dans les recherches des contemporains pour des raisons que nous évoquerons dans la prochaine partie. Il n'est donc pas étonnant que le concept de résolution de problème soit mentionné dans l'introduction explicative de ces rapports. Toutefois, lorsqu'il s'agit d'évoquer les stratégies d'investissements publics, le discours s'oriente alors vers une conception de l'IA comme technologie novatrice permettant de répondre à la demande, c'est-à-dire pouvant adresser des solutions pratiques novatrices efficaces et rentables. La conception de l'intelligence ne semble plus du tout s'apparenter à la résolution de problème dans le sens assigné par Minsky et ses compères : la capacité de mobiliser des stratégies face à des tâches du quotidien et prendre une décision menant à une action en vue d'un objectif.

« Others define AI as a system capable of rationally solving complex problems or taking appropriate actions to achieve its goals in whatever real world circumstances it encounters. » (NSTC, 2016 : 16)

Dans le rapport des États unis, on décerne néanmoins une certaine nuance. On perçoit dans ce « d'autres définissent l'IA » une prise de conscience des multiples définitions assignée à l'IA. On l'a vu, ils ont aussi inséré la définition « grossière » de l'intelligence comme un processus de rétroaction avant d'émettre cette compréhension relative à la résolution de problème.

« Les tâches relevant de l'IA sont parfois très simples pour les humains, comme par exemple reconnaître et localiser les objets dans une image, planifier les mouvements d'un robot pour attraper un objet, ou conduire une voiture. Elles requièrent parfois de la planification complexe, comme pour jouer aux échecs ou au Go. Les tâches les plus compliquées requièrent beaucoup de connaissances et de sens commun, par exemple pour traduire un texte ou conduire un dialogue. » (LeCun, 2016 : 8)

Le chercheur en IA connexionniste assimile ici, lui aussi, l'intelligence par la résolution de problème dont les moyens diffèrent en fonction de la complexité du problème rencontré. Selon sa perception, les chercheurs en IA symbolique s'attelaient à des types de problèmes mettant en jeu une « planification complexe », alors que les chercheurs en IA connexionnistes, eux, s'attellent à un autre type de problème qui ne peut pas être simplement planifié et dont de simples règles assimilées au raisonnement logique ne peuvent appréhender l'ampleur de sa complexité. Ce type de problème requerrait une méthode différente de celle de la déduction.

Selon lui, s'atteler à ce genre de problème nécessite un nombre de connaissances importantes. C'est par la quantité de données (appelées « connaissances ») et dans l'organisation de celles-ci que l'on pourrait résoudre des problèmes de haute complexité. Tout en restant dans le continuum de la résolution de problème, ce chercheur se dissocie de la conception de l'intelligence de l'IA symbolique par la nuance qu'il apporte aux différents degrés de complexité de chaque problème entraînant inévitablement une méthode divergente plus adaptée basée sur l'induction.

De la même manière, le chercheur contemporain Max Tegmark ne se dissocie pas entièrement de la conception de l'intelligence comme résolution de problème tout en apportant une nuance fondamentale représentative de l'évolution de celle-ci.

« Je définis « l'intelligence » de manière très inclusive comme étant notre capacité à atteindre des objectifs complexes, car je veux y inclure les deux intelligences, biologique et artificielle. » (Tegmark, 2018)

Il est évident que la représentation de l'intelligence comme étant la capacité d'une entité à faire face à des problèmes complexes n'a pas entièrement disparu au profit d'une simple optimisation de nos capacités de base. Elle est toujours ancrée dans les représentations sociales des chercheurs lorsqu'ils se soumettent à l'effort de définir cette instance. Une différence profonde émerge toutefois : un objectif doit être atteint, plutôt qu'un problème se doit d'être résolu. *Atteindre* entend parvenir à une limite abstraite. L'accent est dans le chemin pour y parvenir et non pas dans le problème en soi. Nous avons ici affaire à une logique compétitive et marquée. L'idée n'est plus de trouver le résultat, ni même de déterminer la solution à un problème, mais de parvenir à un résultat, d'accomplir un objectif fixé. La logique est mathématique : le problème étant une équation qu'il s'agit de « résoudre ».

Le discours prononcé en 2003 par Minsky lors d'une conférence TED permet de confronter la conception d'un fervent défenseur de l'IA symbolique attaché à la conception de l'intelligence comme résolution de problème dans un contexte où la recherche en IA s'attelle à développer des applications concrètes pratiques dans une optique d'optimisation plutôt qu'à déceler les secrets du fonctionnement cognitifs de l'esprit humain.

« Je pense donc que dans les 20 prochaines années, si nous pouvons nous débarrasser de toutes les approches traditionnelles de l'intelligence artificielle, comme les réseaux neuronaux et algorithmes génétiques et les systèmes à base de règles, et qu'on se contente de viser un peu plus haut pour dire, pouvons-nous faire un système qui puisse utiliser toutes ces choses pour le bon type de problème? Certains problèmes sont bons pour les réseaux neuronaux; nous savons que d'autres, des réseaux neuronaux ne leur servent à rien. » (Minsky, 2003)

Minsky semble être prêt à s'ouvrir aux approches plus modernes en proposant cette idée de combinaisons des méthodes sans pour autant remettre en question sa conception de l'être humain et de son intelligence. Il réalise les limites de l'approche de l'IA symbolique et reconnaît les résultats des méthodes connexionnistes. Mais cela ne suffit pas à le défaire complètement de ses ancrages théoriques puisqu'ils dépendent d'une conception spécifique de l'intelligence ancrée aux confins des représentations propre à l'univers dans lequel ses recherches ont évolué. Remettre en question les fondements des systèmes d'IA qu'il a imaginé revient à remettre en question quelque chose de beaucoup plus profond et d'ancré : son rapport au monde, sa représentation de l'humain, les frontières de ses catégories d'analyse.

À la lumière de chacun de ces discours, il est clair que l'IA est née dans cette conception de l'intelligence comme capacité de résolution de problème. On peut également y constater la manifeste évolution de cette conception vers une optimisation, un stade de l'évolution accompagné d'un délaissement de la recherche théorique basée sur l'étude de l'intelligence dans une ambition idéologique.

## **1.2. L'optimisation**

Si l'expression « résolution de problème » est toujours d'usage dans les discours contemporains, son sens, lui, s'est déplacé. Alors que le mot « problème » représentait pour les chercheurs de l'IA symbolique toute activité quotidienne entreprise par l'individu exigeant la mise en place d'une stratégie, elle évoque pour les chercheurs contemporains les grands maux de la société : l'écologie, la pauvreté, la famine, la science, etc.

« On pourrait dire que l'Intelligence Artificielle (IA) est un ensemble de techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux. » (LeCun, 2016 : 8)

Ainsi, l'intelligence n'est plus par définition la capacité de résoudre des problèmes ; l'intelligence est le moyen de résoudre des problèmes. En d'autres termes, l'intelligence n'est plus une fin, mais un moyen.

« - Et au risque de pas avoir l'air intelligent, c'est quoi la définition de l'intelligence? C'est une excellente question. En gros, quand on prend des décisions et que ces décisions donnent des bons résultats, on prend des bonnes décisions et on utilise notre intelligence. Et pour être intelligent, ça prend des connaissances, et l'intelligence artificielle, c'est un sujet qui consiste à mettre cette intelligence et ces connaissances dans l'ordinateur. » (Bengio, 2017c)

Le passage d'une conception de l'intelligence lié à la capacité de résolution de problème à la capacité d'optimisation ne pourrait être plus évident que dans le discours de Yoshua Bengio. La capacité de résoudre des problèmes est toujours présente, mais réduite à la prise de décision dont le but ultime est d'atteindre de « bons résultats ». L'intelligence sert des finalités différentes : être productif, efficace, générer de l'argent, gagner du temps, faire progresser ; soit optimiser. Ainsi, créer une intelligence artificielle n'est plus une fin en soi, mais le moyen de rendre optimales toutes actions et projets. Tout ce qui est accompli en IA est fait dans l'optique d'atteindre l'état le plus favorable possible : soit de viser une amélioration continue afin d'atteindre le stade le plus proche du perfectionnement permettant une rentabilité notable.

L'optimisation, c'est tout d'abord une question de rentabilité temporelle. La performance de la machine est associée au gain de temps, et donc à la rapidité du processus.

« D'abord, la situation des chabots n'est en effet pas géniale. C'est pour cela qu'on y travaille. Si on avait un système de qualité à ce niveau-là, pour résoudre nos problèmes et nos interrogations rapidement, ça serait quand même très utile. Il y a des aspects de notre vie que l'on pourrait gérer de manière plus efficace avec de bons chabots. Ça, c'est la raison pratique. » (Pineau, 2018)

« En quelques années seulement, on est passés de 20 secondes par image à 20 millisecondes par image. Mille fois plus vite. Comment avons-nous fait ? Avant, les systèmes de détection d'objets prenaient une image similaire et la découpaient en zones, puis un classifieur analysait chacune d'elles, et on considérait qu'il était performant s'il détectait des objets. Mais ça signifiait analyser une image des milliers de fois autant d'évaluations en réseau neuronal pour une détection. On a alors entraîné un réseau unique à réaliser ces détections. Il produit les délimitations et les probabilités de classes simultanément. » (Redmon, 2017)

« Par exemple, pour reconnaître des visages, l'ordinateur n'est pas vraiment meilleur que l'humain. Par contre, si on me donne la photo d'un inconnu et que je dois l'identifier en fouillant dans une base de données qui contient 10 millions de portraits, ça me prendra des années, tandis que l'ordinateur, parce qu'il est plus puissant, va y arriver en quelques secondes. » (Bengio, 2017b)

Comme ces chercheurs le soulignent ici, il s'agit de créer des systèmes pouvant résoudre des problèmes plus rapidement pour des raisons pratiques et dans un souci d'efficacité immédiate. Le concept même d'efficacité est intrinsèquement lié au gain de temps. La satisfaction des résultats ne provient plus de la similarité et de la proximité avec notre intelligence, mais de la rapidité d'exécution des opérations effectuées. La machine se calque au principe cognitif de traitement de l'information dans une optique d'optimisation du temps. Ce rapport au temps n'est pas seulement présent dans les discours des chercheurs contemporains en IA. Les transhumanistes reprennent à leur tour cet argument pour légitimer le développement de l'IA et son utilisation à des fins personnelles :

« The human brain's processing power is difficult to determine precisely, but common estimates range from 1014 instructions per second (IPS) up to 1017 IPS or more. The lower estimate, derived by Carnegie Mellon robotics professor Hans Moravec, is based on the computing power needed to replicate the signal processing performed by the human retina and assumes a significant degree of software optimization. » (Humanity+, 2019)

On retrouve aussi la recherche d'optimisation du temps dans l'optique de transfert de responsabilité à la machine :

« They can think and act for us, for example, clean, move objects, diagnose medical problems, provide care, write software, make the best weather predictions, compose music and play games with us. » (ICTC, 2015)

En transmettant à la machine la capacité de traiter de l'information, résoudre des problèmes, généraliser et prédire, on leur permet d'exécuter des tâches que seuls les humains étaient auparavant capables d'effectuer. Dans la même logique que l'automatisation, le transfert des compétences à la machine permet à l'individu de se concentrer sur d'autres activités moins aliénantes. On peut aussi y voir une façon d'utiliser cette technoscience comme un outil d'assistantat permettant de mobiliser les compétences sociales et cognitives de l'individu sans « perdre de temps » avec des tâches de calculs, d'observations et d'analyses plus redondantes qui demandent un temps de travail plus conséquent. C'est en tous cas l'argument mis en avant par la déclaration de Montréal sur l'IA responsable :

« Il est possible de créer des systèmes autonomes capables d'accomplir des tâches complexes que l'on croyait réservées à l'intelligence naturelle : traiter de grandes quantités d'informations, calculer et prédire, apprendre et adapter ses réponses aux situations changeantes, et reconnaître et classer des objets. » (Déclaration IA responsable, 2019)

Ce transfert de compétence cache cependant un penchant socialement bien moins bénéfique notamment en ce qui concerne l'impact que l'IA va avoir sur le marché du travail. Beaucoup de secteurs de l'emploi vont être touchés par cette prochaine automatisation « cognitive ». C'est d'ailleurs cet aspect qui semble en premier inquiéter la population comme le souligne cet article du journal *Le Monde* :

« Le think tank Institut Sapiens annonçait, le 21 août, que 2,1 millions d'actifs « ont une forte probabilité de voir leur emploi disparaître dans les prochaines années ». Une des premières estimations, faite dès 2013 par deux chercheurs de l'université d'Oxford, l'économiste Carl Benedikt Frey et l'expert en intelligence artificielle Michael A. Osborne, (Humanity+, 2019) que près de la moitié des emplois d'aujourd'hui n'existeraient plus dans vingt ans. » (Le Monde, 2018)

Les transhumanistes sont, eux, relativement optimistes en ce qui concerne cet enjeu de taille. L'Homme, une fois débarrassé de toute aliénation, pourra enfin se concentrer à des tâches

bénéfiques à lui-même et développer ses capacités cognitives au service d'un monde meilleur en progression constante.

« Humans are not simply apes who can obtain more bananas using our intelligence as a tool. Our intelligence enables us to desire new things, such as art, science, and mathematics. If at any point in your indefinitely long life you become bored with the greatly improved world, it may only indicate that the time has come to bump up your intelligence another increment. » (Humanity+, 2019)

Dans cette optique, l'intelligence nous permettrait de désirer d'autre chose, d'aller toujours plus loin et pas seulement grâce au transfert de compétence qui implique une optimisation de temps, d'argent et d'efficacité. Pour les transhumanistes, c'est l'instance même de l'intelligence dont les animaux ne sont pas dotés qui introduit cette habileté de besoins infinis. Le déterminisme du « vouloir plus et mieux » est caractéristique de la vision transhumaniste.

Selon l'expression populaire « Le temps c'est de l'argent », le gain de temps rejoint directement l'optique de la rentabilité économique.

« Les technologies d'amélioration peuvent s'appliquer à beaucoup de domaines. Imaginons que d'ici dix ans on parvienne à développer une molécule sûre et bon marché qui aide les enfants à faire leur cursus plus rapidement, en deux ans de moins par exemple. Voilà une manière d'atteindre le même but, le savoir pour tous, à meilleur marché en termes de dépenses publiques. » (Bostrom, 2008)

Le transhumaniste Nick Bostrom démontre par ces propos qu'il est bien question d'utiliser la technologie à des fins économiques. Comme sa définition l'indique, un investissement est la mise à disposition de ressources financières dans l'optique d'un retour accru de ce capital économique. En investissant dans l'IA, les gouvernements comme les entreprises espèrent arriver à des retombées financières fructueuses. Si ce n'est pas cet argument qui est mis en avant dans les discours destinés à la population civile, c'est pourtant l'argument central qui motive les investissements. En effet, cette analyse permet de rendre compte du fait que le profit est toujours l'argument sous-jacent tous les autres. Pourquoi gagner du temps ? Pour gagner de l'argent. Pourquoi transférer les tâches à la machine ? Pour faire l'économie salariale d'un employé.

« Dites-moi quelle nourriture que vous avez achetée, où elle va, quand elle a besoin d'y être et je vous dirai la plus courte, la plus rapide, la moins chère des routes à prendre. Nous allons économiser, éviter les retards et les perturbations et le but ultime, nous allons sauver des vies. De rien. » (Freeman, 2016)

La chercheuse Freeman utilise l'IA (et plus spécifiquement l'analyse de data) comme moyen de s'attaquer à la famine des pays en sous-développement d'Afrique. Dans la perspective économique dominante, la réciproque se vérifie aussi. Comment sauver des vies ? En faillant aux moyens d'accès et donc en économisant sur les frais. Il s'agit de minimiser les coûts, et de maximiser le rendement sous le critère de l'optimisation. Dans les discours récents centrés sur l'IA, la logique de l'efficacité et de profit domine de loin la recherche de théorisation scientifique que l'on retrouve dans les discours des chercheurs en IA symbolique.

Optimiser, c'est aussi gagner en précision ; soit être statistiquement plus rigoureux. Le taux d'erreurs statistique est plus faible chez la machine, car ses capacités de calcul sont plus puissantes et surtout plus rapides. L'erreur humaine constitue un biais important dont la machine fait l'économie.

« In one recent study, given images of lymph node cells, and asked to determine whether or not the cells contained cancer, an AI-based approach had a 7.5 percent error rate, where a human pathologist had a 3.5 percent error rate; a combined approach, using both AI and human input, lowered the error rate to 0.5 percent, representing an 85 percent reduction in error. » (NSTC, 2016 : 20)

Au vu du rapport sur l'IA établi par le gouvernement des États-Unis, la confiance est générée par l'argument de la précision statistique. Il est pourtant intéressant puisque paradoxale de constater que l'analyse statistique comprend une phase aléatoire importante, c'est pourquoi le terme anglais « accurate » dans le sens « au plus proche de » convient davantage que le terme « précision » qui entend une analyse de détail et d'exhaustivité que ne prend pas du tout en compte l'IA. Au contraire, l'IA connexionniste se base sur la généralisation comme un des moyens d'atteindre l'optimisation. Le processus de généralisation consiste à étendre à l'ensemble l'analyse ou l'interprétation faite d'un phénomène à partir de l'observation de quelques cas seulement. Ce procédé utilisé en IA (et en particulier en apprentissage automatique) se base sur la méthode par induction. À partir d'un échantillon de données, la machine va établir une généralisation à partir de ce qu'elle observe dans un échantillon de données limitées. C'est pourquoi la donnée est une denrée si précieuse de nos jours. Ce n'est pas la donnée en soi qui détient de la valeur, mais sa quantité. Plus le nombre de données est important, plus la généralisation sera précise et juste. Le procédé de la généralisation entraîne une conclusion hâtive d'une valeur théorique pauvre tant elle ne prend en compte tout ce qui se trouve à la marge de la norme. À la base de nombreuses décisions remises la machine, il semble fondamental de mettre en lumière ce procédé. En effet, l'argument mis de l'avant est la fiabilité de la machine en comparaison avec la rationalité biaisée de l'Homme par ses émotions et ses

états d'âme. On peut dire que l'IA est statistiquement parlant plus fiable. Néanmoins, l'être humain guidé par son instinct est conscient des conséquences de ses actes ; ce qui n'est pas le cas du système d'IA. La marge d'erreur a beau être plus faible quantitativement, son importance qualitative n'est pas moins grande. Nouveau prisme de la rationalité, nous misons davantage sur les statistiques que sur la pensée et l'instinct pour nous guider dans nos actions quotidiennes :

« The goal of machine learning is to create a trained model that will generalize—it will be accurate not only on examples in the training set, but also on future cases that it has never seen before. While many of these models can achieve better-than-human performance on narrow tasks such as image labeling, even the best models can fail in unpredictable ways. » (NSTC, 2016 : 19)

La statistique est l'étude d'un phénomène à partir de la collecte de données. Elle s'appuie sur la théorie des probabilités. Du latin *probabilitas*, le terme désigne de son étymologie l'antonyme de certitude. Le fait qu'un phénomène, une proposition, une théorie soit probable n'assure pas qu'il soit vrai.

« C'est intéressant et c'est toute la partie qu'on ne comprend pas, d'une certaine façon. Prenez un ordinateur qui nous donne un algorithme avec un élément aléatoire. Il n'est plus aléatoire quand on connaît le random seed, les conditions initiales de ce système-là. Les ordinateurs sont complètement déterminés. Pour le monde physique, humain, on ne sait pas encore ! Qu'est-ce qui est de l'ordre du déterminisme, qu'est-ce qui est de l'ordre de l'aléatoire ? Plus on comprend, moins il y a d'aléatoire. Et même avec l'aléatoire, on peut chercher du statistique. Si je suis une voiture et que j'ai croisé 5 000 piétons en une semaine et que j'en trouve dix dans l'ensemble qui ont eu un comportement imprévisible, c'est déjà statistique. » (Pineau, 2018)

Cette méthode associée aux mathématiques se dit scientifique, mais est scientifique que ce qui est valide, exhaustif et vérifié (dont l'hypothèse de départ est confirmée). La méthode statistique ne coche aucune de ces cases. C'est sous le prédicat du chiffre et du calcul mathématique trop souvent associé à la scientificité que la pleine confiance lui est accordée. Pour preuve, c'est tout le fonctionnement de l'État moderne qui repose sur les statistiques et les probabilités. Nous remettons à la machine la responsabilité de choisir pour nous sur la base d'une généralisation de ce qui est *probable* et non certifié d'arriver.

La généralisation permet en effet la prédiction : un autre des moyens centraux de l'optimisation. Prédire, c'est prendre de l'avance ; soit gagner du temps et donc de l'argent. Autrement dit, prédire est le moyen ultime d'optimiser. Si les algorithmes d'IA ne sont pas encore mobilisés à cet effet pour l'instant, c'est la visée recherchée vers laquelle on tend.

« L'un des aspects de l'apprentissage prédictif, c'est de vraiment comprendre comment le monde change et d'en concevoir un modèle. Ce n'est pas juste de savoir comment il change

de manière passive, mais, quand j'interviens sur quelque chose, comprendre ce que ça va causer comme changement dans le monde. C'est ça, l'apprentissage prédictif, et c'est une question fondamentale liée à l'IA » (Pineau, 2018)

« What if people could predict natural disasters before they happen? Better protect endangered species ? Or track disease as it spreads, to eliminate it sooner? » (Google, 2019)

Prédire, c'est éviter le pire, pouvoir anticiper et régler les problèmes avant même qu'ils arrivent. Ce à quoi la multinationale Google promet de s'atteler sur leur site web en ouvrant des laboratoires de recherche dans les métropoles du monde.

« Une hypothèse possible est l'apprentissage prédictif. Si l'on entraîne une machine à prédire le futur, elle ne peut y arriver qu'en élaborant une bonne représentation du monde et de ses contraintes physiques. Dans un scénario d'apprentissage prédictif, on montre à la machine un segment de vidéo, et on lui demande de prédire quelques images suivantes. Malheureusement, le futur est impossible à prédire exactement et la machine s'en tient à produire une image floue, une mixture de tous les futurs possibles. » (LeCun, 2016 : 12)

Le chercheur Yann LeCun admet ici les limites de la prédiction : on ne peut prédire exactement le futur. C'est aussi pour ça qu'on se base sur des techniques de généralisation à base de méthode statistique. Il est impossible de prédire ce qui va arriver, on peut toutefois tenter de prédire ce qui va probablement se passer dans un avenir plus ou moins proche et se fier à cette probabilité pour mettre en place des stratégies d'action visant à trouver des solutions aux problèmes du monde. La question financière étant au centre des considérations en IA, on imagine bien que les dépenses ne sont pas mobilisées seulement « au cas où » et donc possiblement en vain. C'est pourquoi la plupart des algorithmes de prédiction ne sont pas aujourd'hui mis à la disposition des gouvernements pour assurer la sécurité du monde, mais à la disposition des entreprises et multinationales.

« In order to foster an environment that supports long-term research, we acknowledge our inability to predict exactly which ideas that we work on today will be the most useful in the years to come. Substantial research freedom combined with a thoughtful and expansive view of how research can produce value at Google allows us to leverage the expertise of our staff to prioritize projects. » (Google, 2018)

L'entreprise Google a compris que la prédiction, au-delà de prévenir des dangers potentiels futurs, sert à déterminer la demande à venir et ainsi devancer les besoins des consommateurs afin de répondre à une demande. Une « demande » contrôlée et déterminée par l'entreprise elle-même de telle manière que les consommateurs croient avoir besoin d'un produit dont l'existence même leur était encore inconnue. La multinationale ne prédit pas ; elle prévoit. Grâce aux algorithmes d'IA (publicités ciblées, proposition de films, de musiques), elle détient le contrôle de nos désirs futurs. Si nous sommes toujours responsables de ce que nous aimons

et de ce que nous n'aimons pas, le contrôle du panel de choix sur lequel on se base est de plus en plus déterminé par l'IA.

« Mon équivalent pour ce message à laisser à nos descendants pour les aider à construire l'intelligence artificielle ou pour les aider à comprendre l'intelligence humaine est la suivante : l'intelligence devrait être perçue comme un processus physique qui tente de maximiser la capacité d'action future et éviter les contraintes de son propre futur. » (Wissner-Gross, 2013)

« Et la réponse, je pense, est oui. [ $F = T \nabla S\tau$ ] Ce que vous voyez est probablement le plus proche équivalent à un  $E = mc^2$  pour l'intelligence que j'ai jamais vu. Ce que vous voyez ici est l'assertion que l'intelligence est une force,  $F$ , qui agit de façon à maximiser la capacité d'action future. Elle agit de façon à maximiser la capacité d'action future, ou encore elle garde le maximum d'options possibles, avec une certaine vigueur puissance  $T$ , avec la diversité de futurs accessibles possibles,  $S$ , jusqu'à un horizon temporel futur,  $\tau$ . En bref, l'intelligence n'aime pas être prise au piège. L'intelligence tente de maximiser la liberté d'action future et garde des portes ouvertes. Étant donné cette unique équation, il est naturel de se demander, que pouvons-nous faire avec ça ? À quel point est-ce utile pour établir des prévisions ? Est-ce qu'elle prédit l'intelligence humaine ? Est-ce qu'elle prédit l'intelligence artificielle ? » (Wissner-Gross, 2013)

Alex Wissner émet ici une réduction directe de l'intelligence à une équation mathématique de prédiction et de maximisation. Ainsi, l'intelligence serait dans cette perspective le processus permettant de maximiser la capacité d'action future. On retrouve une caractéristique si valorisée dans le critère de l'optimisation qu'elle rentre intrinsèquement dans la représentation de l'intelligence humaine. Quelqu'un d'intelligent serait, selon cette conception, quelqu'un qui sait mobiliser les ressources dont il dispose afin d'émettre un résultat optimal. L'intelligence se définirait donc dorénavant en termes instrumentaux dans une logique de production et de contrôle.

### 1.2.1. Données : nouveau modèle de représentation du monde

« Nos cerveaux ne communiquent pas entre eux avec une large bande passante, alors que les ordinateurs, de leur côté, peuvent se connecter entre eux de manière beaucoup plus rapide et se coordonner pour chercher dans les bases de données. » (Bengio, 2017b)

Les données représentent pour Yoshua Bengio les connaissances à partir desquelles on apprend. Plus on en a, plus on gagne en intelligence. Là encore, il est évident que l'optimisation est la nouvelle finalité escomptée. En effet, la quantité de données stockées dans la mémoire d'un programme permet au système d'avoir accès à une base de connaissance beaucoup plus large et immédiate que l'être humain. Grâce à la puissance de calcul qui permet un traitement extrêmement rapide de l'information, la machine permet une optimisation bien plus importante que celle de l'Homme. Selon cette perspective, le niveau d'intelligence est non seulement

déterminé quantitativement (par rapport au nombre de connaissances et non par la qualité et pertinence de celles-ci), mais aussi via la vitesse d'accessibilité à cette base de connaissance. Cette conception de l'intelligence comme procédé d'optimisation permet d'affirmer que l'intelligence de la machine dépasse celle de l'être humain.

« La plupart des connaissances que nous avons comme humain ne sont pas innées. On les acquiert par l'expérience, à travers notre éducation, par ce qu'on observe dans notre quotidien, par l'interaction avec notre environnement. C'est ça, l'idée de l'apprentissage chez les ordinateurs. » (Bengio, 2017b)

« Les méthodes manuelles se sont avérées très difficiles à appliquer pour des tâches en apparence très simples comme la reconnaissance d'objets dans les images ou la reconnaissance vocale. Les données venant du monde réel, les échantillons d'un son ou les pixels d'une image sont complexes, variables et entachés de bruit. » (LeCun, 2016 : 8)

Les données sont perçues comme la représentation de l'environnement. Elles constitueraient en cela le modèle du monde réel. Ce serait par l'observation du monde, en grandissant dans celui-ci et en faisant l'expérience de ces modalités qu'on deviendrait intelligent. Si les données représentent le monde, il suffirait donc que le programme s'imprègne de ces dernières pour qu'il devienne intelligent à son tour.

« La médecine aujourd'hui, elle est faite à partir de très peu de données, quelques dizaines de personnes qui essaient un médicament ou pas. Et là, on va pouvoir utiliser des millions, sinon des milliards de personnes pour nourrir les données qui vont permettre à l'ordinateur de savoir ce qui nous convient le mieux comme traitement et découvrir des nouveaux traitements. » (Bengio, 2017c)

« Est-ce qu'on peut faire confiance à une voiture autonome? Moi, j'attendrais encore un peu avant d'embarquer. J'ai embarqué. Je pense qu'il n'y a pas de danger immédiat, puis dans la conduite ordinaire quand il n'y a pas d'accident et qu'il n'y a pas de danger, ça fonctionne très bien. Mais là où j'ai des doutes, c'est qu'on est pas rendu au point où l'ordinateur qui conduit la voiture va bien réagir dans les situations dangereuses qui sont rares. Et puis comme l'ordinateur apprend à partir d'exemples, il y a «malheureusement» pas assez de gens qui meurent dans des accidents. Il y a pas assez d'exemples pour apprendre. » (Bengio, 2017c)

Au-delà de la notion de données comme modèle le plus représentatif de notre monde, c'est encore la quantité de celles-ci qui prime. Il ne suffit pas d'avoir les données les plus adéquates à la situation, mais d'en avoir un nombre suffisamment important pour que le modèle soit le plus représentatif possible. Il n'y a donc pas de limite à la quantité nécessaire : la logique appliquée ici insinue que plus on a accès à un important nombre de données, plus on s'approchera d'une expérience réaliste du monde dans lequel l'individu évolue. Or, l'Homme n'a pas besoin d'un accès illimité à des connaissances pour mener sa vie et agir à travers le

monde. Il a en revanche besoin de connaissances qualitativement pertinentes à sa propre expérience.

« Puis, j'ai partagé avec lui une séquence d'images d'une maison qui brûlait. Il a regardé les images et il a dit : « C'est une vue incroyable ! C'est spectaculaire ! » Et ça m'a donné des frissons. Il a vu un événement horrible, qui peut bouleverser et détruire une vie, et il pensait que c'était une chose positive. » (Mitchell, 2017)

Cet exemple démontre parfaitement la limite que constitue l'approche essentiellement quantitative des données. L'expérience que l'on a du monde ne se limite pas à un traitement d'information par captements de signaux. Une signification sémantique assignée à ces signaux est essentielle à une compréhension du monde environnant. Plutôt que de remettre en question les assises de son programme, la chercheuse Margaret Mitchell décide de voir en l'apprentissage profond un outil utilitaire avantageux plutôt qu'une simulation de l'intelligence humaine :

« C'est ainsi qu'on pourra commencer à faire évoluer la technologie informatique pour qu'elle soit complémentaire de nos propres expériences » (Mitchell, 2017)

L'apprentissage profond permet, selon elle, une complémentarité quantitative à nos connaissances limitées par notre biologie. Par ailleurs, l'exemple évoqué plus haut évoque une autre limite : le problème du biais.

« J'ai compris qu'il discernait le contraste, les rouges, les jaunes, et il trouvait, sans mauvaise pensée, que ça méritait d'être vu. S'il disait ça, c'était en partie parce que la plupart des images que je lui avais apportées étaient des images positives. C'est parce que les gens ont tendance à partager des images positives quand ils parlent de leurs expériences. Se prend-on en portrait avec son portable à un enterrement ? »(Bengio, 2017c)

Les données censées représenter le monde réel sont partagées et donc sélectionnées par les utilisateurs du web. Ceux-ci procèdent à un tri de ce qui doit être partagé et ce qui ne doit pas l'être. Or, les réseaux sociaux nous dévoilent une image virtuelle et donc aseptisée, altérée, souvent magnifiée et contrôlée de la réalité. C'est pourquoi il est erroné (puisqu'approximatif) d'appréhender les données comme une représentation fidèle du monde environnant.

Cet argument est néanmoins contesté par le chercheur Tom Gruber dans son intervention lors d'une conférence TED. :

« Je ne sais ni quand ni quels facteurs de forme sont en jeu, mais je crois que c'est inéluctable, parce que les facteurs de la réussite de l'IA aujourd'hui, la disponibilité de données exhaustives et les capacités des machines à comprendre ces données, peuvent être appliquées à nos vies. Et ces données sont ici aujourd'hui, disponibles pour nous tous, parce que nous menons des vies numériques, sur nos téléphones et en ligne. » (Gruber, 2017)

Les données seraient selon Gruber bel et bien un modèle de représentation fidèle du monde réel et tant donné que nous vivons à présent des vies numériques. Si nous admettons que les données sont en effet le reflet de notre réalité et que le programme s'imprègne d'elles pour bâtir leur intelligence, on peut alors affirmer que l'IA est criblée de stéréotypes et préjugés en tout genre.

« J'ai compris que, pendant que j'améliorais l'IA, fonction par fonction, jeu de données par jeu de données, je créais des lacunes massives, des trous et des biais cognitifs dans ce que l'IA pouvait comprendre. Et, en faisant cela, je codais toute sorte de préjugés. Des préjugés qui reflètent un point de vue limité, limité à un seul jeu de données. Des partis pris qui reflètent des biais humains, retrouvés dans les données, comme les préjugés et les stéréotypes. » (Mitchell, 2017)

L'IA, en intégrant les données des utilisateurs, intègre ainsi par biais l'injustice du monde dans lequel nous vivons. Pouvons-nous donc nous fier et affirmer qu'elle est plus fiable que l'Homme en raison de sa rationalité opérationnelle ?

La conception de l'intelligence que l'on retrouve dans ces discours se calque et se limite à ce qu'il est possible de transmettre à la machine. Si la plupart des chercheurs savent dorénavant que nous ne découvrirons pas le fonctionnement du système cognitif humain par le biais de la recherche en IA, ils n'hésitent toutefois pas à utiliser cette logique comme d'un procédé de vulgarisation afin de convaincre et faire adhérer son auditoire plus facilement à cette technologie. Dans cette optique, il est plus difficile de s'opposer ou même de remettre en question la méthode et le bien-fondé de celle-ci :

« All of the fields mentioned above use big data analysis and big data does have a critical place in the world of intelligent machines and software. In other words, AI offers technology and methodology and thus enables big data to provide industrial organizations with valuable information for effective decision making » (ICTC, 2015 : 6)

Le rapport canadien sur l'IA tient suffisamment pour acquiescer cette représentation de l'intelligence pour baser sa stratégie politique dessus. En effet, le rapport mentionne ici la volonté d'utiliser les programmes d'IA dans le but de prendre des décisions « efficaces ». En

prenant pour principe absolu une conception construite de l'intelligence, des décisions concernant l'ensemble d'un pays sont prises entraînant des enjeux très concrets.

### 1.2.2. Recherche d'applications et fin de la théorie

Rappelons que la relance de l'IA après un long hiver d'hibernation coïncide avec la découverte de l'apprentissage profond possible grâce à la prolifération des données.

« En 2003, Geoffrey Hinton (de l'Université de Toronto), Yoshua Bengio (de l'Université de Montréal) et moi-même à NYU (l'Université de New York), décidions de démarrer un programme de recherche pour remettre au goût du jour les réseaux neuronaux, et pour améliorer leurs performances afin de raviver l'intérêt de la communauté. » (LeCun, 2016 : 10)

L'IA connexionniste des années 2010 n'a pas été développée sous une ambition d'avancée théorique sur le fonctionnement de l'intelligence humaine au travers de la simulation informatique. L'objectif des chercheurs à travers leur méthode est purement opérationnel : il s'agit « d'améliorer les performances », donc de satisfaire les exigences des investisseurs potentiels, soit émettre des résultats pouvant avoir une utilité pratique (MILA, 2019). Inspiré des réseaux neuronaux artificiel imaginé dans les années 1950, leur approche n'a en effet rien de révolutionnaire.

« Puis la chose importante, c'est que malgré que les machines sont encore très stupides, elles peuvent faire des choses extraordinaires déjà. » (Bengio, 2017c)

La machine, selon Yoshua Bengio, n'a pas à atteindre le niveau de l'intelligence humaine pour détenir une véritable valeur opérationnelle. Si l'objectif à long terme était d'obtenir une intelligence artificielle, dorénavant l'objectif premier est d'émettre des résultats capables de fournir à la société et aux entreprises des outils d'optimisation efficaces.

« Les machines ne peuvent pas penser, elles n'ont pas d'émotions, ne comprennent pas la poésie. Nous ne comprenons pas comment elles fonctionnent. Et alors ? Elles peuvent dès maintenant faire ce pour quoi les hommes sont payés. » (Howard, 2014)

L'éloignement de l'objectif de création d'une créature artificielle à l'image de l'Homme comme imaginaient les pionniers du domaine en créant la discipline de l'IA est ici affirmé. Il s'agit dorénavant de mettre à disposition des systèmes d'IA au service des entreprises et des gouvernements en déployant des applications concrètes capables d'accomplir des tâches qui nécessitaient auparavant d'engager un employé.

« Ce que l'on appelle "intelligence artificielle" depuis 1956 ce sont des techniques mathématiques qui n'ont rien à voir avec l'intelligence. Il n'en reste pas moins que ces techniques (deep learning, machine learning, etc.) sont très intéressantes. Mais la machine ne crée pas, ne réfléchit pas, et les humains conservent pleinement la main sur ces techniques. » (Julia, 2019)

Le pessimisme envers une possible modélisation informatique de l'intelligence humaine est apparent à travers cette dernière citation. Le fait que le projet idéologique semble inatteignable ne les fait en rien reconsidérer le bien fondée de l'IA et en aucun cas ne porte atteinte à son développement. Au contraire, ce fait permet une certaine justification. En concentrant les recherches sur les aspects techniques et les applications pratiques qui peuvent découler de cette poursuite idéologique infructueuse, on ne cherche pas à créer une créature artificielle apeurante pouvant un jour nous dépasser, mais seulement à développer de nouvelles technologies pouvant nous simplifier la vie. Le culte pour la technique étant de nos jours à son comble, cette perspective est à la fois très vendeuse (pour preuve, elle attire les investissements en masse) et réconfortante.

« On a voulu rendre la détection encore plus accessible et pratique. Grâce à la combinaison de l'optimisation de modèles la binarisation du réseau et l'approximation, on peut utiliser la détection d'objets sur un téléphone. Je suis ravi, car nous disposons dorénavant d'une solution puissante pour ce problème de vision artificielle faible. Il peut servir à quiconque pour concevoir ce qu'il veut. Pour le reste, à vous de jouer ainsi qu'à ceux dans le monde qui ont accès à ce logiciel. J'ai hâte de voir ce que les gens feront de cette technologie. » (Redmon, 2017)

La fin de la poursuite idéologique engendre une rupture évidente avec la quête de vérité associée à la recherche fondamentale chère aux humanistes des Lumières. Le discours ci-dessus du chercheur Joseph Redmon démontre très clairement que la recherche opérationnelle dans une quête d'efficacité prime sur le reste.

Les entreprises mettent l'accent sur la quête de profit en stimulant l'innovation. Leur discours entier est tourné vers la création de services utilisant l'IA.

« Pour créer des produits et services d'IA qui bénéficient à tous, tous devraient avoir les mêmes opportunités de contribuer à la recherche en IA. » (MILA, 2019)

« Solve Intelligence. Use it to make the world a better place. » (DeepMind, 2019)

« Make machines intelligent. Improve people's lives » (Google, 2018)

Délaissant tout fondement conceptuel, la recherche en IA n'est plus que centrée sur le développement de biens de consommation. Seul ce qui sert les intérêts corporatifs est valorisé et donc subventionné. L'introduction de Google AI en est la preuve formelle :

« Bringing the benefits of AI to everyone. At Google AI, we're conducting research that advances the state-of-the-art in the field, applying AI to products and to new domains, and developing tools to ensure that everyone can access AI. » (Google, 2019)

L'IA constituerait ainsi un bien de consommation pouvant faciliter la vie de tous, au service du monde. La page web de Google AI se dit informative avant tout dans une visée démocratique. L'optique purement publicitaire de la marque, bien que dissimulée, ne trompe personne : on présente un produit (l'IA) pouvant prendre toutes les formes, participer à toutes tâches, faciliter la vie de tous, répondre aux besoins des consommateurs, quelle que soit leur demande et dans tous les secteurs de l'économie. Il s'agit de montrer que l'IA est une technologie à part, au-dessus des autres dans le sens où elle peut servir toutes sortes d'intérêts et répondre une demande globale et diverse.

« DeepMind is the world leader in artificial intelligence research and its application for positive impact. We're on a scientific mission to push the boundaries of AI, developing programs that can learn to solve any complex problem without needing to be taught how. » (DeepMind, 2019)

« Impact, not just research. This new division gets a placeholder name: X. It has a first project: self-driving cars. Is it a research center? An incubator? Is it philanthropic? No. It's a Moonshot Factory looking 5-10 years into the future and aiming to invent new technologies and launch them into the world. » (X, The Moonshot Factory, 2019)

L'utilisation du terme « impact » démontre que les sociétés Moonshot Factory et DeepMind (toutes les deux rattachées à Google) ne se destinent pas à saisir les mystères du monde, mais à créer des technologies pouvant avoir un réel et concret effet sur le monde. La dévalorisation de la recherche est aussi mise en évidence : « Pas juste de la recherche ». La quête de la vérité est dépossédée de toute valorisation. Il faut créer des technologies, mais aussi de la demande, de l'offre, du besoin, du profit, des solutions et des résultats afin de garantir des conditions d'avenir souhaitables.

Toutefois, nous retrouvons aussi dans certains discours, une tentative d'explication biologique de l'apprentissage profond dans l'optique de résister à la critique de l'éloignement progressivement d'une volonté théorique et donc scientifique.

« Ça conduit à imaginer des structures qui sont aussi hiérarchiques, dans lesquels chaque couche détecte des combinaisons de motifs présents sur la couche précédente. Cette idée de hiérarchie multicouche et de profondeur est donc très naturelle, et on la retrouve dans le cerveau. » (LeCun, 2018 : 39)

« Nous pensons que comprendre l'apprentissage profond ne nous servira pas uniquement à construire des machines plus intelligentes, mais nous aidera également à mieux comprendre l'intelligence humaine et ses mécanismes d'apprentissages. »  
(LeCun, 2016 : 1)

Pour le chercheur LeCun, la justification théorique reste importante. Associée à la scientificité, elle constitue encore un argument d'autorité essentiel à la mise en application pratique de la théorie. Ce type de discours permet de rassurer la population : il n'est pas question de développer une technologie sans raison dont le seul intérêt est économique. Seuls quelques chercheurs s'obstinent encore à démontrer le lien entre les qualités humaines et les compétences de leur système d'IA. Pour convaincre de la pertinence de l'IA et attirer les investissements, les chercheurs de l'IA symbolique n'avaient d'autre choix que de passer par une théorisation basée sur l'étude scientifique de l'être humain et de son intelligence. Aujourd'hui, la pertinence est démontrée par la mise en application. La valorisation du progrès scientifique décline progressivement au détriment de la valorisation du progrès technologique. Nous aurons bientôt plus besoin de justifier scientifiquement pour placer une technologie sur le marché tant le but d'optimisation et de résultats concrets immédiats dépasse de loin l'ambition d'en savoir plus sur notre monde et sur nous-mêmes.

« Si son ambition initiale était d'imiter les processus cognitifs de l'être humain, ses objectifs actuels visent plutôt à mettre au point des automates qui résolvent certains problèmes bien mieux que les humains, par tous les moyens disponibles. » (Villani, 2018 : 2)

Villani est conscient de ce déplacement d'une volonté idéologique de théorisation scientifique du système cognitif humain via son artificialisation vers une recherche empirique essentiellement basée sur la conception d'applications pratiques dont le seul objectif est l'optimisation.

À travers l'analyse de discours, il est clair que pour les chercheurs, les gouvernements et les entreprises, la finalité de l'IA n'est plus dans une recherche de théorisation, ni même d'une modélisation de l'intelligence humaine (recherche fondamentale), mais dans le développement d'applications concrètes, de nouvelles technologies pouvant générer des solutions plus efficaces, plus rapides ; soit d'optimiser. Pour les transhumanistes, en revanche,

l'idée est d'arriver au point de ce qu'il appelle la singularité, soit le moment où l'intelligence artificielle aura dépassé l'intelligence humaine. Ce point de non-retour représente l'accès à ce qui est communément appelé l'IA forte (et donc consciente). Autant l'amélioration recherchée de l'être humain dans l'idéologie transhumaniste se fait par les moyens de la technique (et donc des applications et technologies concrètes), autant l'idéologie se fonde sur cette conviction en une modélisation possible de l'intelligence fondée sur une croyance profonde en le dépassement de l'humain par la machine auquel le seul moyen de réplique sera de se confondre avec la technique. Cette croyance dont les assises proviennent directement du modèle informationnel entend une indifférenciation ontologique entre l'être humain et la machine, permettant de placer sur une même échelle de comparaison ces deux entités. Nous reviendrons sur ce point essentiel dans la seconde partie de cette analyse. Les journalistes, quant à eux, n'entendent pas bien la distinction entre les deux. Pour certains, l'enjeu premier concernerait le dépassement de l'Homme par l'IA en vue d'une perte de contrôle total entraînant ultimement la fin possible de l'humanité. Pour d'autres (selon la direction du journal), l'enjeu est économique et le débat recouvre dans cette optique les infinies promesses technologiques de l'IA.

### 1.2.3. Remise en question de la définition informationnelle de l'intelligence.

Nous venons de démontrer par le biais de l'analyse de discours que la conception de l'intelligence des chercheurs, journalistes, entreprises et transhumanistes lorsqu'ils sont amenés à parler de l'IA se rejoint toutes plus ou moins autour de cette vision informationnelle et opérationnelle. Certains auteurs, journalistes ou même chercheurs remettent toutefois en question cette notion acquise d'intelligence.

« L'intelligence -- qu'est-ce que c'est? Si nous jetons un œil à l'histoire de la façon dont l'intelligence a été vue, un exemple précurseur fut la célèbre citation d'Edsger Dijkstra disant que la question « Une machine peut-elle penser ? » est à peu près aussi intéressante que la question « Un sous-marin peut-il nager ? » Edsger Dijkstra, en l'écrivant, la dirigeait comme une critique des pionniers des sciences informatiques, comme Alan Turing. » (Wissner-Gross, 2013)

Le chercheur Alex Wissner remet ici en cause la pertinence même de se questionner sur la capacité d'intelligence de la machine. Si c'est question lui semble vaine, c'est parce que la visée de l'IA aujourd'hui est opérationnelle. Au temps de l'IA symbolique, l'objectif final résidait dans la réponse à la question « est-ce qu'une machine peut penser ? » Dans l'optique des pionniers de l'IA, si un jour on pourrait affirmer qu'en effet la machine pense, alors ; d'un, on

aura réussi à transférer par modélisation l'intelligence à la machine ; de deux, on aura compris ce qu'était l'intelligence.

« C'est vrai qu'on a franchi un cap à partir de 2007 avec l'apparition du big data du machine learning. Les ordinateurs peuvent maintenant faire mieux que les humains, par exemple, en médecine (pour la radiographique, NDR). Mais la vérité est que les ordinateurs ne font que "reconnaître". Ils n'ont pas de connaissance, ni d'intelligence. » (Julia, 2019)

Luc Julia ne tente pas d'expliquer ici ce qu'est l'intelligence selon lui. Il reconnaît seulement la réduction au fonctionnement purement informationnel et opérationnel de cette intelligence accordée à la machine.

« Quant à l'intelligence artificielle, ça n'existe pas. L'intelligence est une propriété du vivant. L'ordinateur qui a battu le champion du jeu de go est incapable d'aller chercher son pain ou de résoudre un problème de maths. Il y a beaucoup plus d'intelligence dans une abeille que dans un ordinateur. » (Testart, 2018)

Pour ce biologiste de carrière, l'IA n'existe pas puisque l'intelligence est caractéristique du vivant, de la nature. À partir d'une vision dualiste des catégories du monde, il est impossible de concevoir la production artificielle d'une entité appartenant au monde du vivant. Jacques Testart met sous un même continuum les animaux et l'être humain de par leur appartenance au monde des vivants. C'est pourquoi il conçoit en revanche que l'abeille puisse être dotée d'intelligence. La machine faisant partie de la catégorie de la technique, de l'artifice, ne figure pas dans ce même continuum. C'est pourquoi la notion même d'intelligence artificielle n'a absolument aucun sens selon le biologiste. C'est d'ailleurs pour ça que (nous y reviendrons dans le point 3.2) cette discipline n'aurait pu voir le jour sans un renversement du dualisme nature/ artifice.

« On apprend aux intelligences artificielles à penser comme des humains PAS ENCORE Le terme même d'«intelligence artificielle» est en partie responsable de cette erreur. Parler d'intelligence, c'est impliquer qu'un ordinateur puisse être doté de cette faculté cognitive proprement humaine. » (Le Figaro, 2018)

*Le Figaro* reprend ce même argument en répondant négativement à la question : est-ce que l'IA peut penser ? Selon la direction du journal en question, l'être humain et la machine sont séparés d'une frontière infranchissable pour le journal *Le Figaro*. Mais est-ce dans la même optique que le biologiste Jacques Testart (en plaçant l'humain dans la catégorie de la nature, et la machine dans la catégorie de la technique) ; ou est-ce dans une position anthropocentrique intransigeante ? En raison de ses propriétés cognitives uniques, l'Homme se trouverait seul face au non-humain (incluant les animaux et les objets et donc le vivant et la technique). Le lieu de rupture n'est ici pas évident puisque non développé.

« Or le premier danger que présente le développement de l'intelligence artificielle consiste à donner l'illusion que l'on maîtrise l'avenir par le calcul. Réduire la société à des nombres et la gouverner par des procédures algorithmiques est un vieux rêve qui nourrit encore les ambitions humaines. Mais dans les affaires humaines, demain ressemble rarement à aujourd'hui, et les nombres ne disent pas ce qui a une valeur morale ni ce qui est socialement désirable. » (Déclaration IA responsable, 2019)

La déclaration de l'IA responsable de Montréal alerte son auditoire face à cette illusion d'intelligence concédée à la machine. Le bien-fondé de cette technique paraît évident pour son utilité technique tant que la nuance entre l'intelligence humaine et la capacité de calcul de la machine est bien comprise par tous. Ils émettent par-là une critique du culte voué au calcul et mettent en garde de ses implications.

Nous constatons à travers cette analyse de discours que nous ne sommes pas les seuls à adopter une position critique envers le modèle informationnel.

## 2. Le réductionnisme

Maintenant que nous avons illustré par les propos des acteurs de l'IA leur conception de l'intelligence, nous tenterons dans cette section de mettre en avant le réductionnisme par lequel le développement de l'IA est concevable.

« Puis l'idée que plusieurs chercheurs ont depuis plusieurs décennies, c'est que, on peut la comprendre de manière assez simple, avec juste quelques équations. C'est pas sûr, on sait pas si notre intelligence, on pourrait la résumer en quelques équations, mais si c'est le cas, c'est... Bien, c'est extraordinaire. Ce serait comme un peu pour la physique où on a réussi à comprendre le monde avec quelques équations. » (Bengio, 2017c)

L'objectif de l'IA à tous les stades de sa conception est de la rendre réductible à une équation mathématique. C'est seulement ainsi qu'elle pourra être modélisée et donc insérée dans un programme informatique. La croyance d'une réduction possible de cette propriété humaine en une simple équation mathématique est représentative d'une conception informationnelle du monde. On retrouve de plus dans cette citation de Yoshua Bengio une réelle valorisation de la possibilité d'une telle réduction : « ce serait extraordinaire ».

« - Donc, ce que vous voudriez, c'est de trouver l'équation qui résumerait l'intelligence ? C'est ce qu'on essaie de faire ! C'est comme des mathématiques. Mon intelligence peut se résumer avec une équation mathématique ? C'est une hypothèse. On ne sait pas comment fonctionne le cerveau humain, mais je pense (et je ne suis pas le seul !) qu'on peut résumer nos capacités d'intelligence par un petit nombre d'équations. Et, évidemment, une fois qu'on a ces équations, on peut les faire tourner sur des ordinateurs. » (Bengio, 2017a)

En effet, la réduction en IA n'est pas du tout vue comme quelque chose de négatif, qui dépouille l'essence même d'un concept ou d'un phénomène. C'est d'ailleurs le terme « résumé » et non « réduit » que décide d'employer Yoshua Bengio pour convaincre du positivisme relatif à la méthode. La réduction est perçue comme positive justement parce que, par ce dépouillement, elle permet une explication relative de l'objet considéré. On sait que si l'on cherche par la méthode fonctionnelle (de Descartes) à expliquer et examiner chacun des propriétés et mécanismes d'une seule entité, le travail est si exhaustif qu'il en devient vain. Le seul moyen d'expliquer l'intelligence est de le concevoir dans son ensemble et de focaliser sur son processus global. Pourquoi ce serait extraordinaire de pouvoir « résumer » pour ne pas dire « réduire » l'intelligence en une équation mathématique selon le chercheur ? Tout simplement parce que si on peut réduire l'intelligence en une équation mathématique, alors, on peut l'expliquer assez simplement et concrètement pour pouvoir la transférer à la machine et développer des applications pratiques.

« Je vais représenter ces trois choses : les pixels en entrée et les synapses du réseau neuronal, et l'oiseau, la sortie, avec trois variables :  $x$ ,  $w$  et  $y$ . Il y a peut-être un million de  $x$  : un million de pixels par image. Il y a des milliards ou des billions de  $w$  qui représentent le poids de toutes ces synapses du réseau neuronal. Et il y a très peu de  $y$ , de sorties présentes dans le réseau. « Oiseau » n'a que six lettres. Prétendons que la formule est simple :  $x$  « fois »  $w = y$ . Je mets fois entre guillemets, car ce qu'il se passe vraiment est bien sûr une série complexe d'opérations mathématiques. C'est une équation. Il y a trois variables. Nous savons tous qu'avec une équation, vous pouvez trouver une variable si vous connaissez les deux autres. » (Arcas, 2016)

« De cette façon, parlant de tous ces différents sujets et les mettant ensemble, je me suis demandé, il y a quelques années, s'il y avait un mécanisme sous-jacent à l'intelligence que nous pourrions tirer de tous ces différents sujets ? Y a-t-il une équation unique pour l'intelligence ?

Et la réponse, je pense, est oui. [«  $F = T \nabla S\tau$  »] » (Wissner-Gross, 2013)

Les deux chercheurs Blaise Agueri y Arcas et Alex Wissner empruntent la même perspective d'une possible réductibilité de l'intelligence en équation mathématique. Ils n'hésitent pas d'ailleurs à employer la démonstration dans le but de convaincre leur auditoire. Il paraît très clair à travers ces discours que traduire l'intelligence en équation mathématique représente le principe le plus concret, et donc le plus valorisé de l'explication de l'intelligence.

Cette représentation de l'intelligence comme réductible à une équation est aussi représentée dans les discours des entreprises. En effet, pour reprendre le slogan de DeepMind « Our mission is to 'solve intelligence' » (DeepMind, 2019), l'intelligence serait quelque chose qu'il faut résoudre. Or, la notion de « résoudre », en plus d'être associée à la

prise de décision et à la stratégie d'action face à un problème, est indissociable de la notion d'équation : « résoudre une équation ». L'objectif de Deepmind ne se limite pas à résoudre le problème de l'intelligence, mais à résoudre l'équation de l'intelligence qui a été modélisée à cette fin.

« Artificial intelligence (also known as computational intelligence) is defined as “the human-like intelligence exhibited by machines or software”<sup>1</sup>. It is theorized that intelligence of humans can be described and intelligent machines or software can simulate it. » (ICTC, 2015 : 3)

Le discours du gouvernement canadien retranscrit aussi ce réductionnisme. Si l'intelligence humaine peut être décrite, cela signifie qu'elle est modélisable et donc peut se résumer par une équation mathématique pouvant être insérée dans la machine ; ce qu'ils appellent « simuler » l'intelligence.

On retrouve les grands principes du modèle informationnel dans les discours des acteurs de l'IA comme la prise en compte seule du processus global dans une approche comportementale (en opposition à l'approche fonctionnelle de Descartes) :

« Il y a beaucoup de mystère entourant le fonctionnement du cerveau, mais on connaît un certain nombre de principes de fonctionnement qui peuvent nous inspirer. » (Bengio, 2017b)

Nul besoin de comprendre et savoir expliquer la totalité du fonctionnement de l'intelligence, ni même de décrypter chaque composant du cerveau pour s'inspirer des principes de base et développer l'IA. Selon cette approche, seules les interactions, connexions entre éléments sont à prendre en compte dans l'analyse scientifique. Tout se subordonne au principe d'information. Ainsi, la matérialité est négligeable.

« Une machine intelligente n'a pas besoin d'avoir de nez, de bras, de pieds... Elle a juste besoin de pouvoir interagir avec nous, d'avoir accès au monde (par l'intermédiaire de caméras, d'Internet, ou par un autre moyen) pour apprendre comment il fonctionne et pour pouvoir nous aider dans les tâches où on voudrait qu'elle nous aide. » (Bengio, 2017b)

La mise au second plan du support matériel est très représentative du modèle informationnel. Dans cette perspective, où tout système est informationnel et s'explique par le procédé général sans tenir compte de son support, tout s'explique par le même procédé général. En tant que nouvel ordre de représentation, toute théorie scientifique relative à l'étude de l'esprit à partir de la cybernétique s'appuie sur le modèle informationnel.

« Et dans ce programme que j'ai entrepris il y a plusieurs années, j'ai étudié une variété de différents fils conducteurs à travers la science, à travers une multitude de disciplines, qui convergeaient, je le pense, vers un unique, mécanisme sous-jacent pour décrire l'intelligence. » (Wissner-Gross, 2013)

L'interdisciplinarité en tant que source du domaine de l'IA permet une vision globale d'un même phénomène et ainsi une vision unique. Légitimé par cette interdisciplinarité, le monisme informationnel fait disparaître toute autre perspective possible du concept d'intelligence.

## **2.1. Indifférenciation ontologique**

Les fondements du modèle informationnel inhérents de la cybernétique permettent une réduction de l'Homme à un simple système de traitement d'information. Toute entité, aussi bien naturelle qu'artificielle, est alors placée sous ce même ordre de configuration. Cette indifférenciation ontologique profonde, représentative du paradigme dans lequel l'IA a été pensée, se retranscrit dans les discours des protagonistes de la discipline. La comparaison entre l'être humain et la machine est alors possible et justifiée selon cette perspective qui place ces deux entités sur le même continuum.

« Qu'il soit naturel ou artificiel, il y a trois formes principales d'apprentissage. Nous avons déjà parlé de l'apprentissage supervisé. Les deux autres formes sont l'apprentissage par renforcement, et l'apprentissage non supervisé. » (LeCun, 2016 : 11)

Le chercheur parle du concept d'apprentissage comme d'un concept dissocié de son support, de sa provenance. Entité en soi, l'apprentissage (au même titre que l'intelligence) est compris comme un organisme informationnel.

« Je définis « l'intelligence » de manière très inclusive comme étant notre capacité à atteindre des objectifs complexes, car je veux y inclure les deux intelligences, biologique et artificielle. Et je souhaite éviter l'idée absurde du chauvinisme carbonique qui dit, vous ne pouvez être intelligent que si vous êtes fait de viande. Récemment, le pouvoir de l'I.A. s'est développé de façon incroyable. » (Tegmark, 2018)

On retrouve ici aussi cette dissociation de la fonction de sa forme. Que l'intelligence soit naturelle ou artificielle ne joue en rien sur la compréhension du concept. Yoshua Bengio, à l'image de beaucoup d'autres chercheurs, compare l'humain à la machine afin de nous donner un ordre de compréhension de la discipline de l'IA :

« Pas plus qu'on est soumis aux voitures. Et pourtant, elles vont beaucoup plus vite que nous! » (Bengio, 2017b)

On dit qu'où il y a de l'homme il y a de l'hommerie, mais où il y aura des robots, est-ce qu'il y a de la «roboterie»? Est-ce que les robots seront capables du pire? Est-ce qu'ils vont pouvoir être méchants, envieux, mesquins, jaloux? (Bengio, 2017c)

« Mais tu sais, ultimement, même nous autres, quand on dit que c'est bon, c'est ça qu'on fait nous aussi en fait. Tu sais, nous autres, on a l'impression que c'est une réaction spontanée, mais c'est que c'est une réaction qui est provoquée par une analyse très rapide quand même aussi là. » (Bengio, 2017c)

En mettant en parallèle la vitesse de la voiture à la vitesse de l'être humain autant qu'en comparant directement l'IA à l'intelligence humaine, le chercheur tient pour acquis que l'humain et la machine se situent sur la même échelle. À partir du moment où l'on compare deux entités entre elles, c'est que leur nature est considérée du même ordre.

« Nous pourrions devenir plus intelligents. Je pense que nous sommes assez intelligents, comparativement à des chimpanzés, mais nous ne sommes pas assez intelligents pour faire face aux énormes problèmes auxquels nous sommes confrontés, que ce soit en mathématiques abstraites ou pour comprendre les économies ou l'équilibre du monde qui nous entoure. » (Minsky, 2003)

L'intelligence du chimpanzé est comparée à celle de l'être humain qui est comparée à celle de la machine. L'intelligence, en tant que qualité à part, selon qui la détient et en quelle quantité, a la capacité de résoudre des problèmes plus ou moins complexes. Or, si l'on suit la logique de Minsky, les problèmes que constituent les maux de la société engendrés par l'humain ne sont pas maîtrisables par ce dernier, tant leur degré de complexité dépasse leur capacité intellectuelle. Seule la machine semble, selon le chercheur, pouvoir détenir des solutions efficaces. De l'animal à la machine, en passant par l'être humain, le classement hiérarchique du degré d'intelligence selon Minsky est ici évident. Ce qui l'est d'autant plus (et qui rejoint notre propos) c'est la linéarité sur laquelle ces trois entités se suivent.

Dès lors que la différence ne peut se mesurer en termes de degré, on identifie entre ces deux entités une différence de nature. Or, l'intelligence de la machine est souvent représentée comme soit un degré inférieur de la machine, soit à un degré supérieur ; mais quoi qu'il en soit ces deux instances se trouvent sur le même continuum de comparaison.

« On devrait plutôt parler de stupidité artificielle que d'intelligence artificielle, parce que malgré les avancées technologiques, les machines sont encore très stupides. » (Bengio, 2017c)

« Il faut investir dans la recherche fondamentale et on sait pas combien de temps ça va prendre avant d'arriver au niveau d'intelligence des humains. Peut-être 10, 50, 100 ans, on le sait pas. » (Bengio, 2017c)

« Malgré tous ces progrès, nous sommes encore bien loin de produire des machines aussi intelligentes que l'humain, ni même aussi intelligentes qu'un rat. » (LeCun, 2016 : 11)

Cette possibilité de comparaison sous un même continuum prouve qu'elles appartiennent, dans la conception des protagonistes de l'IA, au même ordre de représentation.

« Prenons l'humain le plus intelligent qui ait jamais vécu. Beaucoup de gens penseraient à John von Neumann. L'impression que von Neumann laissait aux gens, y compris les plus brillants mathématiciens et physiciens de son temps, est plutôt bien documentée. Si ne serait-ce que la moitié de ce que l'on dit sur lui est à moitié vraie, il n'y a aucun doute qu'il soit l'un des plus brillants esprits qui aient existé. Considérons le spectre de l'intelligence. Voici John von Neumann. Et nous, nous sommes ici. Et là nous avons un poulet. (Rires) Pardon, un poulet. (Rires) Pas besoin de rendre cette présentation plus déprimante qu'elle ne l'est déjà. (Rires) Il semble néanmoins plus que probable que le spectre de l'intelligence s'étende beaucoup plus loin que nous ne pouvons le concevoir, Si nous construisons des machines plus intelligentes que nous, elles vont très probablement explorer ce spectre, plus que nous ne pouvons l'imaginer, et elles nous surpasseront, plus que nous ne pouvons l'imaginer. » (Harris, 2016)

Sam Harris dans sa présentation orale donnée lors d'une conférence TED illustre son propos en proposant une courbe sur laquelle il place l'intelligence de l'animal, de l'humain et de la machine à des niveaux différents (dans l'ordre croissant). Pour l'animal, il prend l'exemple du poulet, qu'il place très loin derrière l'Homme en termes d'intelligence. Il place sur cette même courbe l'homme qu'il considère comme le plus intelligent du monde. Cet homme n'étant autre que John Von Neumann, créateur du premier ordinateur et pionnier de l'IA, ce qui en dit long sur sa conception de l'intelligence. Il place un espace entre ce scientifique et tous les autres hommes qu'ils considèrent d'une intelligence inférieure dans l'optique d'introduire par la visualisation un écart important entre l'animal et l'Homme dans son degré d'intelligence. Le choix de l'animal en question n'est pas anodin : le poulet représente davantage un produit de consommation aux yeux des Hommes qu'un être sensible et donc pouvant potentiellement être doué d'intelligence. Cette démonstration graphique d'aucune valeur scientifique réelle n'est présentée que dans l'optique d'illustrer la distance de degré de l'intelligence humaine et l'IA qui selon lui est plus importante encore que celle entre le poulet et l'humain.

« C'est presque sûr que l'intelligence artificielle aura des moyens que l'humain n'a pas. » (Minsky, 2003)

« Parce qu'il y a toutes sortes de contraintes, comme la puissance de calcul, qui empêchent que ça se développe sans fin. C'est même raisonnable de penser que la limite supérieure n'est pas très loin au-dessus de l'intelligence humaine. » (Minsky, 2003)

L'IA va-t-elle dépasser l'intelligence humaine et quand ? Ce questionnement (au cœur des discours scientifiques, transhumanistes, éthiques et civils) expose l'adhésion absolue au paradigme cybernétique. Le simple fait de comparer les deux entités montre qu'on les place instinctivement sur une même échelle de valeurs, le degré étant la seule variable de distinction.

Craindre ou espérer que l'IA dépasse l'intelligence humaine démontre la conception alignée des deux entités dans une indifférenciation totale de leur nature.

« C'est qu'on espère qu'un moment donné, on va dépasser le niveau de fiabilité des humains et qu'on va sauver un nombre très, très grand nombre de vies. » (Bengio, 2017c)

Aucun des discours ne met en avant les compétences que l'IA pourrait avoir que l'humain n'a pas, comme pour les autres outils techniques. La hache coupe, l'Homme non. L'avion vole tandis que l'Homme non. La raison réside dans l'indifférenciation de fond, de nature : l'IA n'est pas fiable tandis que l'humain est compétent ; l'IA est *plus* fiable que l'Homme.

« En nous couchant un soir, l'intelligence artificielle (celle des ordinateurs interconnectés) serait légèrement moins puissante que la nôtre, puis le lendemain matin, nous serions placés devant une intelligence radicalement supérieure à nous. Je crois que cela est possible, qu'à un certain moment dans l'avenir, il pourrait y avoir cette phase de progrès foudroyant. » (Bostrom, 2004)

« Donc la seule alternative, et c'est ce qui se trouve derrière la porte n°2, est que nous continuions d'améliorer l'intelligence de nos machines année après année. À un moment, nous allons construire des machines plus intelligentes que nous, et ces machines plus intelligentes que nous vont commencer à s'améliorer elles-mêmes. Il y a le risque, théorisé par le mathématicien I.J. Good, d'une « explosion d'intelligence », un processus qui pourrait nous échapper. » (Harris, 2016)

What is a superintelligence ? « A superintelligent intellect (a superintelligence, sometimes called “ultraintelligence”) is one that has the capacity to radically outperform the best human brains in practically every field, including scientific creativity, general wisdom, and social skills. » (Humanity+, 2019)

Les plus convaincus du devancement de l'IA sur l'intelligence humaine sont les transhumanistes. Tous les maux des Hommes peuvent se résoudre selon eux via la technique : la dévalorisation de l'humain et le culte voué à la technique soulignent cette croyance en une supériorité de la technologie sur l'organisme défaillant et limité de l'humain. La différence de degré qui marque l'écart ambivalent entre l'intelligence humaine et l'IA se trouve dans ce principe de limite. Le progrès technique est illimité tandis que l'Homme dans sa composition naturelle se dégrade avec le temps et ne peut tendre naturellement à dépasser les limites biologiques qui lui sont imposées de sa naissance à sa mort. C'est dans cette optique que le dépassement de l'intelligence humaine par l'IA est, selon eux, incontestablement inévitable. Si on reste dans cette conception réductrice moniste de l'intelligence, il est probable que l'on affirme un jour que l'IA a dépassé l'intelligence humaine. Ce ne sera pas forcément le cas pour autant. La machine aura seulement des compétences que l'être humain n'aura pas : la capacité d'un calcul et d'un traitement d'information extrêmement rapide.

## 2.2. Les qualités non modélisables intrinsèques à l'humain

Si le chercheur Blaise Aguera y Arcas conçoit toute propriété comme modélisables, d'autres auteurs argumenteront au contraire que certaines qualités restent et resteront le propre de l'Homme. Elles constituent pour cette raison une limite infranchissable à la reproduction artificielle de l'intelligence.

« Darwin dit que l'intelligence, c'est se comporter intelligemment. C'est ce qui permet de tenir le coup, d'aller de l'avant. Pour ça, il faut avoir peur, par exemple. Sinon, en cas de danger, on meurt. Une machine n'a pas peur. Il faut de l'intuition, de la créativité, prendre des risques : tout ça ne peut pas être programmé. Seule la logique est programmable. C'est comme la porte du garage qui s'ouvre toute seule ou le chauffage de votre salon qui se met en route avant votre retour : ce n'est pas une maison intelligente, c'est une maison programmée. » (Testart, 2018)

Pour Jacques Testart, une machine ne peut pas être intelligente justement parce que selon sa conception provenant de la définition de Darwin, les composants de l'intelligence humaine ne sont tout simplement pas modélisables. L'IA n'est qu'opérationnelle, elle permet le développement seul d'une série d'applications nous servant d'outils censés nous simplifier la vie de l'ordre de toutes les autres technologies. C'est ainsi qu'il conteste l'assignation d'intelligence donnée à cette technologie.

« Mais est-ce que ça correspond à notre intuition de ce qui est réellement créatif ? Est-ce suffisant pour créer quelque chose d'original et beau, ou est-ce que la créativité nécessite une intention et une conscience de ce qui est en train d'être créé ? Peut-être que dans ce cas la créativité vient réellement des programmeurs, même s'ils ne comprennent pas vraiment le processus. De toute façon, qu'est-ce que la créativité humaine ? Est-ce plus qu'un système de neurones interconnectés développés par le biais d'un processus biologique algorithmique et par les expériences aléatoires qui forgent ce que nous sommes ? Ordre et chaos, machine et humain. Voilà les moteurs qui animent les projets de créativité synthétique qui créent en ce moment musique, sculptures, peintures, poésie et plus encore. On pourra toujours se demander s'il est justifié de juger ces créations comme créatives. » (Weinberg, 2015)

Le chercheur Gill Weinberg remet en question la question de la possibilité d'une créativité chez la machine. Il conteste avant tout la définition assignée à cette instance en IA comme d'une réaction psychique composée d'un ensemble de signaux qui s'activent dans notre cerveau dont le procédé général est suffisamment simple et observable scientifiquement pour pouvoir le modéliser et ainsi le programmer au sein d'un système artificiel. Il se demande : peut-on vraiment considérer certains programmes d'IA comme créatifs ? Tout dépend de la conception de la créativité. Tout comme l'intelligence, la créativité peut être appréhendée dans une approche constructiviste comme un concept construit.

À ces questionnements, il répond :

« Mais si une œuvre peut vous faire pleurer, ou vous étonner, ou vous faire frissonner de plaisir, est-ce vraiment important de savoir qui ou quoi l'a créée ? »(Weinberg, 2015)

Il ne servirait à rien de savoir d'où provient ce qui a été créé tant ce qui en ressort provoque une émotion bien réelle chez les êtres humains. Cette conclusion est très représentative de la visée purement opérationnelle du développement de l'IA de nos jours. Pour résumer la pensée de ce chercheur : non, la machine n'est pas douée de créativité (partie intégrante de l'intelligence) propre à l'Homme et donc non modélisable, mais si ce qui est développé en IA que les chercheurs attribuent à la créativité provoque la même sortie que la réelle créativité humaine, alors, on peut affirmer que l'objectif est atteint. Dans une optique d'efficacité et d'optimisation, ce qui est recherché en IA aujourd'hui n'est pas de comprendre le fonctionnement de l'esprit humain, mais de développer des programmes étant capables des mêmes résultats que l'humain.

### **3. Entre nature et artifice**

#### **3.1. Créature artificielle à l'image de l'Homme**

« À titre personnel, je crois qu'il y a dans la tradition judéo-chrétienne l'idée que l'homme égalerait Dieu s'il fabriquait un être animé à son image. D'un côté cela signifie qu'il ab céderait à une certaine forme de perfection en produisant une image animée de lui-même puisqu'il s'égalerait à Dieu, ce qui est positif ; mais, d'un autre côté, il y a là une forme de transgression qui paraît maudite. Et cela conduit tout naturellement à craindre que la transgression se reproduise : Dieu a créé l'homme qui le domine, de même que l'homme créant un robot risque de voir ce dernier le dominer... » (Ganascia, 2018)

En voulant fabriquer une créature artificielle à son image, l'Homme souhaiterait avant tout égaler Dieu selon le philosophe et chercheur en IA Ganascia. Le désir de création de cet être se trouve au juste milieu entre la tentation d'égaliser ce qui y a de plus haut, au centre et à la source de tout (Dieu) et la crainte d'enfreindre par là une transgression irréversible. La position de l'IA dans cet entre-deux est soulignée par les nombreux discours en cet égard : tantôt évocateur d'espérance, tantôt évocateur de la fin du monde ; toujours dans la fascination.

Toutes créatures artificielles (de Pygmalion à l'IA) ont été conçues à l'image de l'Homme, dont la visée idéologique est de retranscrire la représentation que l'on se fait de lui et, ultimement, de mieux saisir cette dernière. Pourquoi vouloir alors reproduire l'intelligence de l'être humain ? Simplement parce que, selon la définition qui lui est assignée, c'est ce qui fait sa spécificité, qui le place à part ou pour certains en haut de la hiérarchie des vivants.

« Ce qui fait de nous des humains, c'est un ensemble de choses et notre intelligence. Elle est au cœur non seulement de notre identité, mais aussi de nos forces. » (Bengio, 2017b)

Yoshua Bengio perçoit l'intelligence comme la composante centrale et fondatrice de notre identité en tant qu'être humain. Cela démontre que l'intelligence est comprise dans cette optique comme un avantage sur les autres espèces, comme une instance qui lui est propre. On retrouve cette même perception dans le discours de son collègue Yan LeCun :

« Il est très difficile de nous imaginer une entité intelligente qui n'ait pas toutes les qualités et les défauts des humains, car l'humain est notre seul exemple d'entité intelligente. Comme tous les animaux, les humains ont des pulsions et des instincts gravés dans notre cerveau reptilien par l'évolution pour la survie de l'espèce. Nous avons l'instinct de préservation, nous pouvons devenir violents lorsque nous sommes menacés, nous désirons l'accès aux ressources pour ne pas mourir de faim, ce qui peut nous rendre jaloux. Nos instincts d'animaux sociaux nous conduisent aussi à rechercher la compagnie d'autres humains. Mais les machines intelligentes n'auront aucune raison de posséder ces pulsions et instincts. Pour qu'elles les aient, il faudrait que leurs concepteurs les construisent explicitement. » (LeCun, 2016 : 12)

L'intelligence comme instance est détachée du reste de ce qu'on partage avec les autres animaux selon le chercheur. C'est la spécificité de l'humain qui est reproduite davantage que l'individu dans tous ses attributs. Pour qu'une entité soit intelligente, ses désirs, pulsions et défauts sont négligeables, puisque les animaux, eux aussi, sont dotés de ces mêmes propriétés. Or, ils ne sont pas considérés comme intelligents. Au sein d'autres représentations sociales, ce sont pourtant des qualités intrinsèques à l'intelligence. Le fondement de l'ambition de fabriquer une créature à l'image de l'homme s'est toujours trouvé dans une tentative de compréhension de notre être, une sorte de rétrospective mêlée d'auto-idolatrie. Avec l'IA, la compréhension se fait à travers la modélisation : s'il est possible de modéliser un phénomène alors ce phénomène est considéré comme compris. Il est alors envisageable de le reproduire artificiellement.

« L'autre raison, c'est que le langage fait énormément partie de notre humanité. Ça nous différencie des autres espèces animales, même si certaines ont leurs propres systèmes de communication. C'est fondamental à notre humanité et à nos interactions sociales. Comme chercheuse en intelligence artificielle, si je comprends le langage, je comprends tellement plus ce qu'il se passe dans le monde. » (Pineau, 2018)

Cette citation est à entendre ainsi : si j'arrive à modéliser le langage, alors je modélise l'humain ou du moins ce qui fait de l'humain une espèce à part. Ainsi, je peux appréhender le monde. Cette vision ne peut être plus anthropocentrique : si je comprends l'être humain, je comprends le monde.

Toutefois, nous l'avons vu, l'ambition des chercheurs contemporains s'éloigne progressivement de celle de créer une entité artificielle à l'image de l'Homme.

« Même si un jour on construit des systèmes par certains aspects plus complexes ou performants que les humains, ils vont être construits pour des tâches spécifiques. On associe trop souvent l'intelligence artificielle aux qualités et aux défauts humains. » (LeCun, 2016 : 2)

L'ambition se limite désormais à développer de nouveaux outils permettant une productivité optimale. C'est pourquoi Yan LeCun dit ici vouloir dissocier le domaine de l'IA des qualités et défauts de l'être humain. Il est en effet courant de retrouver dans les discours autour de l'IA une comparaison entre l'Homme et la machine. Ce réflexe de comparaison s'explique, on l'a vu, par l'indifférenciation ontologique instaurée par la cybernétique, mais provient de l'assise idéologique de l'IA : fabriquer une créature artificielle à l'image de l'Homme. Autrement dit, l'IA est née de cette ambition de reproduction de l'être humain et l'on doit sa concrétisation au modèle informationnel.

Aujourd'hui, l'intelligence n'est plus qu'une simple ressource :

« Et, vu la valeur de l'intelligence -- l'intelligence est la source de tout ce qui compte ou alors elle doit protéger tout ce qui compte. C'est notre plus importante ressource. » (Harris, 2016)

L'explication de la nature de l'Homme ne se trouve plus dans la compréhension de son intelligence. Cette dernière reste toutefois la valeur considérée comme la plus importante de l'être humain, car elle constituerait la ressource la plus précieuse qu'il détient. Dans un utilitarisme réducteur, l'intelligence en tant que ressource constitue un simple moyen, un outil dont l'être humain dispose.

### **3.2. Renversement du dualisme**

« En revanche, ce qui s'efface est notre croyance en l'existence d'un « propre de l'homme » ; des facultés que rien – animal ou machine – ne puisse partager avec nous. Et l'effacement de cette croyance nous transforme profondément : depuis cinq siècles, notre morale est fondée – pour le meilleur et pour le pire – sur ce postulat humaniste d'une singularité de l'espèce humaine. Nous devons désormais inventer une autre morale qui s'en affranchit. Une première étape est sans doute de faire l'inventaire de toutes les actions que l'humanisme justifie et qui nous semblent aujourd'hui bonnes ou mauvaises. » (Dowek, 2018 : 250)

L'effacement de la croyance en « un propre de l'Homme », pour reprendre les termes de Gilles Dowek, est dû au bouleversement des catégories qui dirigeaient auparavant notre

vision dualiste du monde. L'humain vs le non-humain, la nature vs la technique : de nos jours, les catégories se dissolvent pour se confondre. Pour de nombreux auteurs, chercheurs et autres, le brouillage des frontières est le seul moyen de sortie d'un anthropocentrisme restreignant. Si d'un côté l'inclusion du non-humain dans nos analyses nous semble légitime et bénéfique à une vision plus juste du monde, elle permet de l'autre côté de justifier le développement sans limites des technosciences comme l'IA.

« Combining computer science with physics and biology to create breakthroughs that help the world. Computer science and natural science are complementary: breakthroughs in one can lead to remarkable advances in the other. The goal of the Applied Science Organization at Google is to cross-fertilize these two fields. » (Google, 2018)

L'IA n'aurait pu être pensée en dehors de ce contexte de renversement de la distinction entre nature et artifice. C'est d'ailleurs l'un des arguments marketing phares de la multinationale Google : c'est seulement dans une complémentarité des sciences naturelles et de l'informatique que de grands progrès seront réalisés.

« Il faut souhaiter que notre cercle éthique s'agrandisse aux animaux et aux formes de vie transhumaines. Mais cela voudra dire aussi qu'on abandonne le concept étroit d'Homo sapiens. » (Bostrom, 2008)

Pour le transhumanisme Nick Bostrom, le concept même d'être humain, ou comme le philosophe l'appelle : « au concept étroit d'Homo Sapiens », est trop limité et pas assez inclusif ; il faudrait donc s'en débarrasser. C'est seulement ainsi qu'il serait possible d'améliorer l'humain techniquement tout en respectant les considérations éthiques. Celles-ci constituent un frein aux ambitions transhumanistes. Puisque le débat éthique est exclusivement tourné vers l'Homme, l'élargissement du concept et la prise en compte des « autres formes de vie » seraient nécessaires pour que l'artificialisation de nos corps biologiques soit enfin envisagée et éthiquement acceptée. Pour Nick Bostrom, si l'idée d'être humain augmenté n'est pas encore adoptée par la majorité, c'est avant tout à cause des catégories d'analyses par lesquels se créer notre rapport au monde. En redéfinissant les concepts majeurs, on peut légitimement intégrer toutes formes de vie en hybridation avec la technique.

« Cela dit, le concept même d'espèce n'est sans doute plus adapté pour comprendre notre avenir. Biologiquement, l'espèce est définie par la faculté qu'ont les individus d'un groupe à se reproduire entre eux, mais pas avec ceux d'un autre groupe. Or la technologie, en combinant le matériel génétique de plantes, d'animaux et d'humains, permettra de franchir ces barrières. À mesure que le matériau devient plus malléable, l'idée d'espèce fixe devient problématique et le critère de reproduction perd son sens. » (Bostrom, 2008)

Tandis que certains (le philosophe Gilles Dowek par exemple) argumenteront que le dualisme ne permet plus de décrire la réalité du monde dans lequel nous vivons, d'autres (Nick Bostrom par exemple) trouvent la dualité des catégories tout simplement problématique. Elle est perçue comme une contrainte, une limitation nous empêchant d'aller plus loin dans l'expérimentation d'une hybridation de la nature et de l'artifice dans le but de s'autoaugmenter. En tant que fondateur de l'association transhumaniste Humanity +, la perspective de Nick Bostrom se retrouve évidemment dans la charte des transhumanistes :

« The human is a biological animal, which evolved approximately 200,000 years ago as the subspecies *Homo sapiens sapiens* (modern humans). The Western world's consensus on what is "normal" for a human biology, life span, intelligence and psychology established certain precedents. Outside these precedents would mean that a human is subnormal or beyond normal. A person who is afflicted with a physical affliction, a mental condition, or degenerative disease would be considered to be outside the normal range. Likewise, a person who has increased physiological performance or cognitive abilities, or lives beyond the human maximum lifespan of 122-123 years, would be considered outside the normal range. This determination of "normal" has not kept up with the advances in technology or science. » (Humanity+, 2019c)

Les transhumanistes évoquent ici la volonté de déconstruire le dualisme et les catégories strictes sous l'argument qu'elles nous restreignent et nous empêchent d'approfondir les frontières de l'Homme limité par sa biologie. La frontière du normal, désigné par le dualisme, doit, selon eux, être abolie.

Au travers des discours centrés sur l'IA, nous verrons que la déconstruction du dualisme est déjà entamée. Pour certains, elle est déjà ancrée dans leurs représentations sociales et l'IA en est la preuve.

### 3.2.1. Artificialisation de la nature

L'appropriation technique de la nature ne pourrait être acceptée sous un dualisme strict. Ainsi, le brouillage des frontières légitime l'artificialisation de la nature. L'IA (comme son nom l'indique) consiste en effet à rendre artificiel ce qui, jusqu'ici, était par définition de l'ordre de la nature.

« Et les chercheurs, depuis des décennies, essaient d'en comprendre la nature. Une manière d'y arriver, c'est d'essayer de la reproduire... Un peu comme on a fait avec les avions : on s'est inspirés des oiseaux. » (Bengio, 2017b)

Comme le mentionne ici le chercheur Yoshua Bengio, c'est en s'inspirant de la nature, et en tentant de la reproduire artificiellement que l'on serait capable de la comprendre.

« Nous sommes confrontés au déclin cognitif. À mesure que nous vieillissons, ce déclin est inévitable. Il se déroule pour certains de manière douce, pour d'autres de façon plus grave, avec l'apparition de la démence, ou bien de manière extrême, avec les maladies neurodégénératives. Notre but est de régler ces problèmes, mais que se passe-t-il si vous voulez restaurer ce qui est perdu ? Il se trouve que la technologie peut apporter des solutions : par exemple, si vous implantez quelque chose dans l'aire de Wernicke [la fonction de cette aire est connue par les lésions qui y sont associées et qui provoquent un déficit de compréhension du langage, oral comme écrit, ndlr], vous êtes susceptible d'améliorer votre capacité de langage, etc... » (Howard, 2014)

Pour le chercheur Howard, non seulement il est question de s'inspirer du fonctionnement biologique de l'être humain par le biais de la neurologie pour concevoir des systèmes d'IA, mais il s'agit surtout d'introduire la possibilité d'une implantation d'artifice au sein même de notre matière biologique. D'après lui, cela permettrait d'apporter des « solutions » aux défaillances que constitue notre nature de par son caractère limité. Il est donc explicitement question d'artificialiser ce qui relève de la nature.

« Posthumans could be completely synthetic artificial intelligences, or they could be enhanced uploads [see “What is uploading?”], or they could be the result of making many smaller but cumulatively profound augmentations to a biological human. The latter alternative would probably require either the redesign of the human organism using advanced nanotechnology or its radical enhancement using some combination of technologies such as genetic engineering, psychopharmacology, anti-aging therapies, neural interfaces, advanced information management tools, memory enhancing drugs, wearable computers, and cognitive techniques. » (Humanity+, 2019c)

Il devient clair que le processus d'artificialisation de la nature constitue le moyen premier pour les transhumanistes d'atteindre leur objectif final : transcender la nature humaine en augmentant techniquement l'humain. Dans cette optique, le posthumain constitue ainsi l'incarnation de l'Homme artificialisé.

Ce processus d'artificialisation de la nature ne pourrait avoir lieu sans une conception mécanique de celle-ci fonctionnant par un processus prédictible et répétitif et donc observable. Rendre artificiel ce qui est naturel revient alors à une simple et permise reproduction mécanique.

« Cependant, si vous prenez du recul et pensez à ce qu'étaient les innovations les plus puissantes qui nous ont permis de construire des machines artificielles sachant nager et des machines artificielles sachant voler, vous trouvez que ce fut uniquement à travers la compréhension des mécanismes physiques sous-jacents de la nage et du vol que nous avons pu construire ces machines. Donc, il y a plusieurs années, j'ai entrepris un programme pour essayer de comprendre les mécanismes physiques fondamentaux sous-jacents à l'intelligence. » (Wissner-Gross, 2013)

La saisie des mécanismes physiques de la nature est nécessaire à sa reproduction artificielle. Cette compréhension doit néanmoins se faire à l'intérieur du modèle informationnel : si toute entité naturelle n'est pas considérée comme un processus mécanique il n'est alors pas envisageable de modéliser les phénomènes de la nature et, de ce fait, de les reproduire artificiellement. Cette manière de concevoir tout état biologique comme mécanique est fondateur de la pensée transhumaniste :

« But if we're doing new things, learning, growing more intelligent, and we still aren't happy, for no better reason than that our cognitive architecture is badly designed, then perhaps it is time to redesign it. » (Humanity+, 2019c)

Le système cognitif est perçu comme une machine dont l'architecture doit être revue afin d'être améliorée.

« Et nous savons que la matière peut produire « l'intelligence générale », la capacité de réfléchir de manière croisée, parce que nos cerveaux y sont parvenus, n'est-ce pas ? Après tout, il n'y a que des atomes là-dedans, et tant que nous continuerons à construire des systèmes d'atomes qui feront montre de plus en plus d'intelligence, nous parviendrons, à moins d'être interrompus, nous parviendrons à implanter une « intelligence générale » au cœur de nos machines. » (Harris, 2016)

L'intelligence humaine ne serait qu'un processus physique concret fait d'atomes. Il n'y a donc, selon cette perspective, aucune raison de penser qu'il n'est pas possible de reconfigurer ce processus physique à l'intérieur d'un support matériel artificiel. Si la plupart des critiques de l'IA se tournent vers le débat de ce qui est modélisable et ce qui ne l'est pas, l'approche constructiviste nous invite à reconnaître que tout est modélisable par le processus d'artificialisation de la nature. Tout dépend de la conception théorique qui guide notre vision du monde.

### *La foi en la technique*

La conception de l'être humain comme une machine se comprend par la dépréciation de tout ce qui a trait au naturel et donc au biologique. La foi en l'être humain et en sa capacité de réflexivité permettant une remise en question des principes et valeurs porteuses de nos sociétés accompagné d'une action politique a, pour cette raison, été substituée par une foi véritable en la technique.

Les transhumanistes sont les porteurs les plus vigoureux de ce culte voué à la technique. Dans leur section FAQ du site web officiel de l'association Humanity+, la question est ainsi posée : « Won't it be boring to live forever in a perfect world ? », à laquelle les transhumanistes répondent avec audace et présomption : « Why not try it and see ? » (Humanity+, 2019c) Ils se

disent ainsi prêts à mettre en application leur projet de vie éternelle et d'humain augmenté et sont disposés à faire face aux implications que cela engendre *a posteriori* tant leur confiance en la capacité de la technique à résoudre tous les problèmes auxquels est confronté l'être humain est absolue.

« Nous avons des problèmes que nous devons absolument résoudre. Nous voulons vaincre les maladies comme Alzheimer ou le cancer. Nous voulons comprendre l'économie. Nous voulons améliorer le climat. Donc nous allons continuer, si nous le pouvons. Le train a déjà quitté la gare et il n'y a pas de frein. » (Harris, 2016)

Perçue comme unique moyen pouvant résoudre les maux de la société, la technique est souvent représentée dans les discours tournés vers l'IA comme la solution permettant de déresponsabiliser l'humain tout en l'élevant vers de nouvelles possibilités. Pour résoudre les problèmes les plus fondamentaux de la société, seule la technique (et donc l'IA) semble pouvoir nous sauver.

« Ne peut-on pas atteindre le « meilleur » que vous évoquez par le truchement de la philosophie, du savoir, de l'éducation? Avez-vous vraiment besoin de la technologie? Oh oui! inévitablement. Le « meilleur » qu'on peut atteindre par la connaissance comporte d'importantes limites. Par exemple, on ne peut certainement pas mettre fin au vieillissement simplement en réfléchissant à cette question. Il vous est loisible, par exemple, de vous convaincre qu'en définitive, c'est une bonne chose de mourir. Mais le résultat, c'est que vous allez mourir quand même après 70 ou 80 ans, ou même avant. Si vous voulez vraiment découvrir le type de maturité ou de sagesse à laquelle on peut aspirer après quelques centaines d'années d'existence, vous ne pouvez pas éviter de passer par la biochimie. Il n'est pas suffisant de « réfléchir » à ce que votre pensée serait après deux cents ans de vie. Et en ce qui a trait à la pensée elle-même, tout comme certains concepts humains sont inaccessibles aux chimpanzés, aux chiens ou aux lapins, parce que leur petit cerveau ne suffit pas à la tâche, de même, nous devons supposer qu'il y a des concepts, des idées et des pensées qui sont tout simplement trop complexes pour notre cerveau humain. » (Bostrom, 2004)

La philosophie des Lumières nous aura appris à réfléchir sur les enjeux de notre monde comme moyen d'y faire face. La philosophie nous amène à contester des acquis et en créer de nouveaux selon des valeurs toujours plus altruistes, plus justes, plus humanistes. Le problème selon le transhumaniste Nick Bostrom, c'est que si la réflexion philosophique transforme en effet notre vision de voir la nature humaine et la société, elle n'apporte toutefois aucun changement radical concret, matériel et instantané de la nature humaine ou de la société comme la technologie peut le faire. Ainsi, dans le contexte d'une société dirigée par le critère de l'optimisation, n'est valorisé que ce qui est observable et apporte des résultats opérationnels palpables. Convaincus que l'IA peut apporter des solutions concrètes et sauver l'Homme des maux de la société, les problèmes éthiques concernant les enjeux engendrés par l'IA s'effacent derrière un sentiment de fierté à participer au « sauvetage » de l'humanité :

« Maintenant, si le niveau de la mer atteint effectivement l'I.A.G., les progrès futurs de l'I.A. seront alors gérés principalement non plus par les hommes, mais par l'I.A., ce qui pourrait impliquer de nouveaux progrès bien plus rapides que les années de recherches nécessaires à l'homme pour aboutir. » (Tegmark, 2018)

Les entreprises utilisent ce même argument pour convaincre les clientèles du bien-fondé de leur investissement en la recherche en IA :

« Applying AI to some of the world's biggest challenges. Through research, engineering, and initiatives to build the AI ecosystem, we're working to use AI to address societal challenges. » (Google, 2019)

« If we're successful, we believe this will be one of the most important and widely beneficial scientific advances ever made, increasing our capacity to understand the mysteries of the universe and to tackle some of our most pressing real-world challenges. From climate change to the need for radically improved healthcare, too many problems suffer from painfully slow progress, their complexity overwhelming our ability to find solutions. With AI as a multiplier for human ingenuity, those solutions will come into reach. » (DeepMind, 2019)

« We create radical new technologies to solve some of the world's hardest problems » (X, The Moonshot Factory, 2019)

Selon leurs discours, ce serait avant tout pour le bien commun, une société meilleure et plus égalitaire, un monde moins pollué, que l'argent est investi dans cette direction. Il paraît toutefois suspicieux que de grandes multinationales cherchant par tous les moyens à contourner leurs obligations fiscales espèrent en tant que retour d'investissement le simple bénéfice d'une société plus juste<sup>5</sup>. C'est dans cette exacte perspective que les gouvernements empruntent le même discours :

« L'intelligence artificielle constitue un progrès scientifique et technologique majeur qui peut engendrer des bénéfices sociaux considérables en améliorant les conditions de vie, la santé et la justice, en créant de la richesse, en renforçant la sécurité publique ou en maîtrisant l'impact des activités humaines sur l'environnement et le climat. Les machines intelligentes ne se contentent pas de mieux calculer que les êtres humains, elles peuvent interagir avec les êtres sensibles, leur tenir compagnie et s'occuper d'eux. » Déclaration de l'IA responsable

« Applications of AI for Public Good  
One area of great optimism about AI and machine learning is their potential to improve people's lives by helping to solve some of the world's greatest challenges and inefficiencies. » AI-US

---

<sup>5</sup> De nombreux articles de journaux relatent de l'évasion fiscale opérée par les multinationales. Parmi eux, l'article de France Info : « Evasion fiscale : "Peut-on reprocher à Google de profiter des lacunes de la loi ?" met en lumière cette réalité et questionne sa légitimité. (France Info et Nicolas, 2014)

La confiance envers l'IA est telle que les gouvernements lui remettent volontiers entre les mains le sort de la société, le devenir de l'humanité ainsi que la santé de la planète. Cette loyauté envers la technique provient d'une perte de confiance envers l'être humain :

« Oui, j'ai toujours fondamentalement confiance dans la technologie. Je pense même que la technologie peut combattre des fléaux comme les *fake news*. On va voir apparaître des agents intelligents capables de vérifier les informations. Il y a toujours une partie sombre dans la technologie, mais c'est parce que les humains en font une mauvaise utilisation. »  
Luc Julia

Régler les problèmes de l'Homme par les moyens de la technique c'est admettre que l'on a perdu toute foi en la possibilité d'une amélioration sociale par et pour les Hommes. La logique est la suivante : puisque ce sont les Hommes qui se sont mis dans cette situation chaotique, pourquoi compter sur eux pour nous en sortir.

« Donc l'option appréciée de certains de mes collègues en I.A., c'est de créer une super intelligence, mais sous le contrôle des hommes, comme un dieu réduit en esclavage, déconnecté de l'Internet, et utilisé pour créer une technologie et une richesse inimaginables au profit de quiconque la contrôle. Mais Lord Acton nous a prévenus que le pouvoir corrompt, et qu'un pouvoir absolu corrompt absolument. Il est donc à craindre que les humains ne soient pas assez intelligents, ou plutôt doués de sagesse suffisante pour gérer un tel pouvoir. » (Tegmark, 2018)

Plutôt que de remettre en question notre usage des ressources de la planète pour répondre à des envies qu'on a transformées en besoin, ce qui impliquerait de transformer non seulement nos modes de vie, mais notre vision du monde ; il est plus facile et rassurant de trouver en la technique la solution à tous nos problèmes.

« Nous avons fait un sondage récemment, et j'ai été surpris de voir que la plupart des gens veulent la création d'une super intelligence, d'une I.A. largement supérieure à nous sur tous les plans. Et presque tous étaient d'accord que nous devrions être ambitieux et essayer de répandre la vie dans l'univers. Mais il y avait moins de consensus quant à qui ou quoi devrait être en charge, et j'ai été plutôt amusé de constater que quelques-uns voulaient que ce soit seulement les machines. » (Tegmark, 2018)

La déresponsabilisation et désolidarisation permet l'éloignement nécessaire à nos consciences pour continuer de consommer sans culpabiliser. Cette perception est bien évidemment portée par les protagonistes dont les discours sont décuplés dans cette analyse : le chercheur, mais surtout les entreprises et les gouvernements dont les intérêts sont tournés avant tout vers la croissance économique. La confiance envers la technique comme solution aux maux du monde constitue moins une réelle conviction, qu'un moyen de persuasion que nous empruntons la bonne voie, voir l'unique voie possible.

Les discours journalistiques diffèrent des autres, justement parce que leurs intérêts ne sont pas d'ordre économique, mais de l'ordre de la divulgation de l'information. En effet, parmi les discours analysés, le discours journalistique est le seul qui met en garde sur la pleine confiance en l'algorithme soi-disant objectif de l'IA :

« L'intelligence artificielle est plus objective que les humains quand elle prend des décisions. FAUX - Les amis suggérés sur un réseau social comme Facebook, l'attribution d'un poste dans l'administration, l'itinéraire d'un chauffeur Uber ou encore les recommandations de films sur Netflix... Les algorithmes sont au cœur de décisions importantes comme quotidiennes. Mais les exemples de discriminations algorithmiques se multiplient. » (Le Figaro, 2018)

L'apport bénéfique de certains progrès techniques et scientifiques est indéniable. C'est la volonté de s'adonner entièrement à la technique comme seul espoir de l'humanité en délaissant la réflexion et l'action politique que nous remettons en revanche en question. Ce culte adonné à la technique provient aussi d'une dévalorisation du biologique et tout ce qui a trait au biologique.

« Et enfin, l'humain ne se trouve pas au sommet de l'intelligence, nous en sommes même très loin. Et c'est vraiment le point crucial. C'est ce qui rend notre situation si précaire et ce qui fait que notre compréhension des risques n'est pas fiable. » (Harris, 2016)

En plaçant l'être humain sur le même continuum des techniques mécaniques et informatiques, on le compare à celle-ci. Or, l'être humain placé sur cette lignée est une machine peu performante, en grande partie en raison de ses limitations imposées par la biologie.

« If the human brain has a "set point" of happiness to which it returns, maybe this is a design flaw and should be fixed – one of those things that we will end up defining as human, but not humane » (Humanity+, 2019c)

Dans l'idéal transhumaniste, les défaillances du corps et de l'esprit seraient réparables au même titre que n'importe quelle machine.

« En fait, nous vivons un moment remarquable dans l'histoire de l'humanité où, guidés par notre refus d'accepter les limites de nos corps et esprit, nous construisons des machines d'une complexité et d'une grâce exquis et magnifiques qui étendent l'expérience humaine bien au-delà de notre imagination » (Booch, 2016)

Le biologique entendrait des barrières infranchissables. Ce sont précisément ces limites qui sont dévalorisées plus que le biologique en lui-même dont on s'inspire par fascination. «Don't limit your challenges, Challenge your limits » Humanity+. Ce slogan officiel de l'association transhumaniste nous invite à dépasser les limites biologiques qui nous restreignent dans une condition fixe et empêche l'augmentation de notre espèce.

« In other words, we want people to be better than well. This is the goal of transhumanism. »  
(Humanity+, 2019a)

Repousser les limites de notre nature pour ne pas se contenter d'être bien, mais « mieux que bien » : voilà le projet transhumaniste. Cette citation renvoie directement au décalage entre les acteurs de ce mouvement qui font partie des hommes les plus fortunés du monde et le reste de la population. Il ne suffit pas de rendre accessible l'IA au plus grand nombre pour combler cet écart, le problème de fond ne résidant pas dans les moyens financiers. En effet, « being well », soit « être bien » serait suffisant pour une grande majorité de la population encore loin du simple état de « bien-être ». La mise en place des conditions nécessaires au bien-être de tous ne devrait-elle pas passer avant l'ambition de dépasser l'état de bien être ? Une poignée de personnes seulement peuvent aspirer à être « mieux que bien ». L'aspiration à vouloir toujours plus dépasse l'aspiration à ce que le reste de la population aille tout aussi bien.

### 3.2.2. Naturalisation de la technique

Selon le processus de naturalisation de la technique, l'IA serait perçue comme entité naturelle. Cette appréhension de la réalité suffirait à justifier et légitimer l'hybridation du corps avec la matière artificielle :

« - Isn't this tampering with nature ? » « - Absolutely, and it is nothing to be ashamed of. It is often right to tamper with nature. One could say that manipulating nature is an important part of what civilization and human intelligence is all about; we have been doing it since the invention of the wheel. » (Humanity+, 2019c)

Selon la logique transhumaniste, puisqu'on doit notre civilisation à la manipulation de la nature, tempérer avec la nature ne peut qu'être bénéfique. À une époque où la population commence à prendre conscience que cette altération massive de la nature est la cause principale de la crise écologique, la glorification de celle-ci par les transhumanistes démontre le détachement face à tout ce qui ne concerne pas les besoins et envies de l'être humain. Alors que le mouvement veut en finir avec tout dualisme qui ne permet pas à l'être humain de s'étendre, leur idéologie reste anthropocentrique et de ce fait, exclut toute considération pour les autres espèces vivantes et pour tout ce qui relève de la nature en général. Le renversement des catégories dualistes n'est souhaitable que pour servir l'ambition d'augmentation de l'humain par la technique. Il n'est en aucun cas question d'appréhender toute entité sur une même lignée horizontale. La suite de cette citation démontre très clairement la logique empruntée par le mouvement pour légitimer leur volonté :

« Alternatively, one could say that since we are part of nature, everything we do and create is in a sense natural too. » (Humanity+, 2019c)

Le processus d'artificialisation de la nature est légitimé par celui de la naturalisation de la technique. Puisque nous sommes des êtres appartenant à la nature, tout ce que nous créons est aussi de l'ordre de la nature. L'IA serait alors une entité naturelle au même titre que tout autre objet technique. En passant dans la catégorie de la nature, l'objet technique peut légitimement être utilisé afin d'altérer la nature, soit artificialiser cette dernière.

« Changing nature for the better is a noble and glorious thing for humans to do. » (Humanity+, 2019c)

Selon cette logique, manipuler la nature ne pose aucun problème épistémologique ; au contraire. Pourquoi vouloir remettre en question une technologie qui peut aider et permettre une avancée sociale et humaine ? Cette question revient souvent dans les débats. Une des réponses pourrait être parce que ces assises ne s'appuient pas sur des valeurs humanistes altruistes et socialement orientées et très souvent sa visée réelle attendue non plus.

« Elles annoncent les effets sismiques, les effets terrestres des mathématiques que nous construisons. Et le paysage a toujours été façonné par cette collaboration étrange et malaisée entre la nature et l'homme. Mais maintenant il y a une troisième force co-évolutionnaire : les algorithmes, le Jongleur, le Carnaval. Et nous allons devoir les comprendre comme faisant partie de la nature. Et dans un sens, c'est vrai. » (Slavin, 2011)

Faut-il inclure les algorithmes d'IA dans la catégorie de la nature, tant la nature a été altérée par la technique ? Selon Kevin Slavin, une reconfiguration de nos catégories est en effet nécessaire dans une optique de représentation plus réaliste du monde technicisé que nous avons façonné. Autrement dit, les catégories dualistes ne suffisent plus à appréhender notre réalité. Le déterminisme technologique conduit à considérer l'IA comme une avancée naturelle autonome, se développant comme stade inévitable de l'évolution de l'humanité. Dans cette perspective, l'Homme est indissociable de la technique ; c'est ce qui le distinguerait des autres animaux. L'être humain ne partant de rien met en place ses propres moyens d'élévation via la technique. Le déterminisme technologique est omniprésent dans les discours des chercheurs en IA : on ne peut arrêter le cours naturel des choses et l'avancée du progrès technologique amenant à la création prochaine d'une IA :

« Étant donné l'importance de l'intelligence et de l'automatisation, nous allons continuer à améliorer notre technologie si nous le pouvons. Qu'est-ce qui nous en empêcherait ? Une guerre nucléaire mondiale ? Une pandémie globale ? Un impact d'astéroïde ? Justin Bieber président des États-Unis ? (Rires) Il faudrait que la civilisation telle que nous la connaissons soit détruite. Il faudrait vraiment imaginer quelque chose de terrible pour que nous arrêtons

de développer notre technologie totalement, génération après génération. Par définition, ce serait la pire chose qui serait jamais arrivée à l'humanité. » (Harris, 2016)

Mettre un terme au développement de la technologie reviendrait à décider d'aller à l'encontre de l'évolution et donc de l'humanité. Plus qu'un simple outil pouvant nous aider à résoudre les problèmes du monde, la technique représente l'unique direction que l'individu peut envisager. Toute autre voie acheminée entraînerait la fin de la civilisation.

« D'ici 20 ans, nous aurons des nano-robots, car une des nouvelles tendances à forte évolution est la miniaturisation de la technologie. Ils entreront dans notre cerveau à travers nos vaisseaux capillaires et connecteront simplement notre néocortex à un néocortex synthétique dans le cloud, nous en fournissant ainsi une extension. À l'heure actuelle, vous avez un ordinateur dans votre téléphone, mais si vous avez besoin de 10 000 ordinateurs pour quelques secondes pour faire une recherche complexe, vous pouvez y avoir accès dans le cloud. Dans les années 2030, si vous avez besoin d'une extension de néocortex, vous pourrez vous y connecter via le cloud directement depuis votre cerveau. Imaginons que je me balade et je me dis, « Oh, je vais croiser Chris Anderson, il vient vers moi. Il faut que je trouve quelque chose d'intelligent à dire... J'ai trois secondes pour ça. Les 300 millions de modules de mon néocortex ne seront pas suffisants. J'en ai besoin d'un milliard de plus. » Je serai alors capable d'y accéder via le cloud. Alors nous disposerons d'un système de pensée hybride fonctionnant sur des composants biologiques et non biologiques, mais la partie non-biologique est sujette à ma loi du retour accéléré. Elle va grandir de manière exponentielle. » (Kurzweil, 2014)

Utilisant le futur de l'indicatif, celui qui est communément appelé le pape du transhumanisme retranscrit parfaitement ce déterminisme. Plus qu'une simple éventualité, l'intégration de l'IA à notre support biologique est un évènement qui selon lui et ses prédictions sur l'avenir, se passera quoiqu'on en décide, dise ou pense.

Par ailleurs, le processus de naturalisation de la technique est légitimé par la contrainte : nous aurions mis en place une société si technicisée, que nous n'aurions d'autres choix à présent que de se transformer soi-même en vue d'une augmentation de nos capacités physiques et cognitives afin de rivaliser avec la technique.

« Réfléchissez un peu que le moyen le plus sûr, recommandé, serait de brancher cette technologie directement sur nos cerveaux. Cela peut sembler être la meilleure option, mais généralement, on essaie d'être sûr de son coup avant de s'enfoncer quelque chose dans le cerveau. Le vrai problème est que simplement concevoir une IA super intelligente semble plus facile que de concevoir une IA super intelligente tout en maîtrisant les neurosciences pour connecter cette IA à nos cerveaux. Si l'on tient compte du fait que les entreprises et gouvernements se sentent probablement en concurrence, et, puisque remporter cette course revient à conquérir le monde, pourvu que vous ne le détruisiez pas juste après, il semble probable que l'on commence par le plus facile. » (Harris, 2016)

L'idée de l'implantation de l'IA dans le cerveau est concevable par le processus de la naturalisation de la technique via l'effondrement du dualisme nature/ artifice. Pour être à la

hauteur de la technique que nous avons créée, la seule option serait de s'adapter nous-mêmes en tant qu'être biologique à la technique et de nous hybrider en son sein afin de devenir des êtres bioniques, ou, comme certains l'appellent : un posthumain.

"If we can effectively merge with A.I. by improving the neural link between the cortex and your digital extension of yourself - which already exists, it just has a bandwidth issue - then effectively you become an A.I. human symbiote. And if that then is widespread, [where] anyone who wants it can have it, then we solve the control problem as well. We don't have to worry about some evil dictator A.I., because we are the A.I. collectively. That seems to be the best outcome I can think of." (Musk, 2019)

Elon Musk est une des figures phares exprimant son apeurement face à la révolution de l'IA. Face à ce qu'il appelle « our biggest existential threat » ou encore : « the demon », sa solution est de l'incorporer directement dans nos cerveaux. Autrement dit, on ne peut pas contrer le démon, mais on peut devenir le démon et ainsi garder le contrôle de ses actions.

Puisqu'il n'est pas envisageable dans cette optique d'aller à l'encontre de l'évolution naturelle de l'IA, il nous reste en revanche la possibilité de débattre des enjeux de l'IA dans le but de minimiser les risques et maximiser les avantages. Conscients des impacts négatifs que l'IA apportera, il s'agit dorénavant d'élaborer des stratégies d'actions pour rediriger ces effets néfastes afin de laisser place aux bénéfiques.

### *Réflexion centrée sur les enjeux*

C'est pourquoi la réflexion de chacun des discours en IA (quel que soit son locuteur) est centrée sur les enjeux :

« Mais comme toute technologie puissante, l'IA peut être utilisée au bénéfice de l'humanité entière ou au bénéfice d'un petit nombre aux dépens du plus grand nombre. L'émergence de l'AI va sans doute déplacer des métiers. Mais elle va aussi sauver des vies (par la sécurité routière et la médecine). Elle va très probablement s'accompagner d'une croissance de la production de richesses par habitant. La question pour les instances dirigeantes est comment distribuer ces nouvelles richesses, et comment former les travailleurs déplacés aux nouveaux métiers créés par le progrès technologique. C'est une question politique et non technologique. C'est une question qui n'est pas nouvelle : l'effet du progrès technologique sur le marché du travail existe depuis la révolution industrielle. L'émergence de l'IA n'est qu'un symptôme de l'accélération du progrès technologique. » (LeCun, 2016 : 13)

« La majorité des chercheurs prévoit l'arrivée de l'I.A.G. d'ici quelques décennies, et si nous plongeons là-dedans sans préparation, ça pourrait être la pire des erreurs de toute l'Histoire de l'humanité. Regardons les choses en face. Elle rend possible une dictature brutale à l'échelle mondiale, des inégalités, une surveillance et une souffrance sans précédent, voire même l'extinction de l'espèce humaine. Par contre, si nous manœuvrons prudemment, nous pourrions avoir un futur fantastique où tout le monde est mieux loti - les pauvres plus

riches et les riches plus riches - et où tout le monde est bonne santé et libre de poursuivre ses rêves. » (Tegmark, 2018)

La visée dans la maximisation des bénéfices et la minimisation des risques n'est pas d'exacerber l'écart de richesses ou d'entrevoir une société plus égalitaire. Si le pouvoir d'achat augmente autant pour les riches que pour les pauvres, alors, l'écart de richesse reste inchangé et les inégalités perdurent. La solution qui semble mettre tout le monde d'accord est de poursuivre dans la marche du progrès technologique et de la croissance économique illimitée tout en gardant le contrôle sur son évolution :

« L'IA peut devenir plein de choses différentes. Mais dans ce cas, ce ne sera pas un véhicule autonome sans destination. Ce sera une voiture qu'on pilote. On choisira quand accélérer et quand ralentir. On décidera de faire un virage ou non. On sélectionnera ce à quoi ressemblera l'IA dans le futur. Il y a beaucoup de choses que l'intelligence artificielle peut devenir. Elle a un énorme potentiel. Et c'est à nous, aujourd'hui, de trouver ce qu'on doit mettre en place pour s'assurer que les effets de l'intelligence artificielle soient bénéfiques pour nous tous. » (Mitchell, 2017)

Seule la recherche des bénéfices est envisagée, tant que ceux-ci dépassent les coûts engendrés (quels qu'ils soient) :

« Our motivation in all we do is to maximize the positive and transformative impact of AI. We believe that AI should ultimately belong to the world, in order to benefit the many and not the few, and we'll continue to research, publish and implement our work to that end. » (DeepMind, 2019)

La déclaration responsable de l'IA à Montréal a été écrite dans cette optique :

« L'intelligence artificielle constitue un progrès scientifique et technologique majeur qui peut engendrer des bénéfices sociaux considérables en améliorant les conditions de vie, la santé et la justice, en créant de la richesse, en renforçant la sécurité publique ou en maîtrisant l'impact des activités humaines sur l'environnement et le climat. [...] Le développement de l'intelligence artificielle présente cependant des défis éthiques et des risques sociaux majeurs. En effet, les machines intelligentes peuvent contraindre les choix des individus et des groupes, abaisser la qualité de vie, bouleverser l'organisation du travail et le marché de l'emploi, influencer la vie politique, entrer en tension avec les droits fondamentaux, exacerber les inégalités économiques et sociales, et affecter les écosystèmes, l'environnement et le climat. Bien qu'il n'y ait pas de progrès scientifique ni de vie sociale sans risque, il appartient aux citoyens de déterminer les finalités morales et politiques qui donnent un sens aux risques encourus dans un monde incertain. » (Déclaration IA responsable, 2019)

Sur la base de l'illusion qu'on ne peut mettre en place une technologie sans que son objectif de départ soit détourné, il convient de le faire de manière responsable afin de jouir pleinement des bénéfices qu'elle peut apporter à la société, tout en se préparant aux conséquences néfastes qu'elle entraînera inévitablement par la même occasion.

Les discours du chercheur Yoshua Bengio sont toujours orientés sur les enjeux éthiques, sociaux, politiques et économiques de l'IA :

« Plusieurs milliers de scientifiques de l'intelligence artificielle ont signé une lettre pour demander aux gouvernements de bannir l'utilisation de l'intelligence artificielle dans les armes létales qui tuent sans l'intervention humaine. Il y a déjà un comité en place à l'ONU qui discute avec différents États, dont le gouvernement américain et le Pentagone. Ça, c'est un enjeu qui nous préoccupe. La peur de perdre sa job et de se faire remplacer par des robots, c'est une peur réaliste. On peut aussi imaginer des gouvernements contrôler leur population avec ces technologies. On ne voudrait pas qu'un gars comme Trump ait accès à des outils lui permettant de traquer les gens et de punir ceux qu'il n'aime pas. Ça serait l'enfer. Même chose pour les entreprises. On ne voudrait pas donner des outils puissants à celles qui feraient n'importe quoi pour leur profit au détriment de l'humanité. On ne souhaite pas non plus les mettre entre les mains de fous ou de terroristes. La peur de perdre sa job et de se faire remplacer par des robots, c'est une peur réaliste. Ça ne va pas se faire demain matin, mais graduellement, des emplois seront perdus à cause de l'arrivée des robots. C'est le cas avec l'automatisation, et si on prend l'exemple d'une usine de voitures, on voit bien qu'un robot qui fabrique plus de voitures pour moins chères, ça nous enrichit... si on met de côté les conséquences environnementales. Une intelligence améliorée va augmenter la richesse. Après, la question, c'est : comment cette richesse sera-t-elle répartie? On ne peut pas laisser la loi du marché, la loi de la jungle, décider comment les choses vont évoluer. Je pense que les gouvernements doivent s'entendre et que les populations doivent réfléchir à la question pour que la répartition soit la plus juste possible. C'est mon opinion, mais elle est quand même assez partagée parmi les scientifiques. » (Bengio, 2017b)

La question se pose : est-ce que le chercheur en IA est le mieux placé pour entrevoir les enjeux engendrés par les technologies que lui-même développe ? Si la participation des chercheurs à la mise en place d'une IA responsable est préférable, il n'est pas pour autant souhaitable que cette politique éthique soit établie par l'ingénieur lui-même en raison des conflits d'intérêts évidents.

L'intégration aux discours de la notion du choix induite par le concept de liberté rassure sur les enjeux dont les médias ne cessent d'évoquer en plus de légitimer le développement de l'IA à l'aveugle.

« We favour allowing individuals wide personal choice over how they enable their lives. This includes use of techniques that may be developed to assist memory, concentration, and mental energy; life extension therapies; reproductive choice technologies; cryonics procedures; and many other possible human modifications and enhancement technologies. » (Humanity+, 2019b)

Selon le principe de liberté propre aux sociétés néolibérales, toute personne aura le choix de jouir ou non des bénéfices de cette technologie. Le développement technologique est ainsi légitimité par la liberté du choix individuel. Pour les transhumanistes, l'individu n'aura qu'à décider s'il veut ou non augmenter son système cognitif au moyen de puces d'IA implantées

directement dans le cerveau. Mais comme Jacques Testart le mentionne, cette question de choix sur la base d'un système libéral est belle sur le papier, mais dessert une tout autre réalité :

« Voyez comme le smartphone a transformé notre manière d'être. On nous dit qu'on a le choix de ne pas en avoir : c'est faux. Essayez, et vous vous retrouverez en marge de la société». Demain, refuser de se faire implanter une puce sous la peau pour surveiller notre état de santé ou ne pas vouloir sélectionner génétiquement son en-décapants que de ne pas posséder de smartphone? » (Testart, 2018)

La question du choix n'est qu'un leurre : a le choix celui qui a les moyens d'avoir le choix. De plus, quand toute une société a adhéré à une technologie, celui qui n'y adhère pas se retrouve ultimement à la marge. Le choix par conséquent n'est pas celui de posséder ou non la technologie de l'IA, mais celui d'être socialement accepté en se confondant à la norme ou décider de résider à la marge de la société.

#### **4. Considérations finales en vue des résultats de l'analyse discursive.**

L'analyse de discours a permis de confirmer notre postulat théorique central : les discours sur l'IA des acteurs en lien plus ou moins direct avec la discipline sont révélateurs d'une représentation réductrice informationnelle de l'intelligence de l'être humain. À travers la partie théorique de notre recherche, nous avons vu que la représentation cachée de l'intelligence en IA se trouve dans la finalité poursuivie du système. Puisque l'IA (à la base) est une tentative de reproduction artificielle de l'être humain, regarder à quelle fin la machine est conçue pour appréhender la conception du fonctionnement de l'intelligence humaine relève de la logique. En d'autres termes, les chercheurs façonnent leur algorithme d'IA dans l'optique que la sortie du programme se trouve au plus proche de leur représentation spécifique de l'intelligence humaine puisque leur création se veut à l'origine bio-inspirée. Plus qu'une simple progression de la définition, l'analyse de discours dévoile qu'une véritable fracture eut lieu. Aujourd'hui, la recherche d'un rapprochement avec la nature n'est plus d'actualité. Dans la recherche de l'optimisation comme finalité convoitée, l'ambition d'une reproduction artificielle de l'intelligence ne trouve plus sa place. Il n'est plus possible d'affirmer que l'IA est une discipline scientifique dont l'objectif rejoint celui des sciences de l'esprit, c'est-à-dire découvrir et expliquer les confins de la nature de l'intelligence humaine. L'IA est devenue un produit technologique au même titre que tout autre objet technique servant la puissance économique par sa capacité de maximiser du temps et donc de l'argent. La recherche de vérité a disparu au détriment d'une pure recherche pratique d'application, délaissant l'ambition idéologique à la base de l'IA : celle de façonner une créature artificielle à l'image de l'Homme.

Toutefois, la représentation de l'intelligence se trouve toujours d'une certaine façon dans la sortie anticipée du système d'IA puisque la capacité d'optimisation constitue dorénavant dans les sociétés contemporaines occidentales la qualité la plus valorisée de l'individu. Ces sociétés productivistes fonctionnent par la contribution active du citoyen à la croissance économique. Ainsi, le rôle de l'individu se limite à être optimal dans chacune des tâches qu'il lui est confié (le travail), comportements (la consommation), et décisions (au quotidien et dans son rapport aux autres et au monde) qu'il entreprend. La dimension de l'optimisation est si importante, qu'elle en devient identitaire. La réduction de l'humain à un simple processus informationnel est encore vraie, mais elle est désormais poussée à son paroxysme : l'Homme dans la représentation sociale des acteurs de l'IA devient une machine à optimiser de l'information. Ceci dévoile une réalité plus profonde : en tant que simple rouage de la grande machine économique, l'être humain est réduit à sa fonction, à ce qu'il apporte, soit à sa « sortie ». En effet, dans cette configuration, il n'est autre qu'une machine complexe. Dans une indifférenciation ontologique propre au modèle informationnel, l'être humain est placé sur la même échelle que la machine dont seule la place qu'il occupe sur cette lignée le différencie de cette dernière. L'analyse de discours a non seulement confirmé ce point théorique, mais l'a rendu indéniable : la comparaison entre l'intelligence humaine et l'IA se retrouve dans chacun des discours. En tant que simple dispositif matériel dont l'individu est seul à pouvoir jouir de sa possession comme bon lui semble, l'être humain peut légitimement être artificialisé. L'appréhension à exercer des modifications volontaires dans le but de s'augmenter succombe alors peu à peu puisqu'elles ne représentent plus une transformation identitaire, mais une mise à jour de la version initiale. Dans cette perspective, la limite constitue une contrainte et la technologie, elle seule, permet de dépasser la frontière infranchissable que représente dorénavant la dimension du biologique. L'IA détient, selon ces représentations sociales, une performance optimale supérieure à l'intelligence humaine. En effet, l'être humain ne peut traiter de l'information toute la journée sans repos, il ne peut se souvenir de tout ce qu'il a appris. Quand bien même il aurait cette capacité de mémoire et de stockage, il ne pourrait y avoir accès aussi rapidement que le fait la machine. Le système n'a pas de revendications salariales, ne demande pas de congé, ne tombe pas enceinte, il est « meilleur » simplement parce qu'il permet de maximiser chaque seconde, fait donc économiser une somme d'argent conséquente ce qui, ultimement permet un profit notable. Sous ce critère ultime de l'optimisation, on peut dire que l'IA dépasse l'intelligence humaine simplement parce qu'elle contribue davantage à la croissance économique que le travailleur consommateur auquel est réduit l'être humain.

Par une naturalisation du progrès technique, cette réalité est difficilement décelable et encore moins contestable. Ce déterminisme cultive l'imaginaire collectif selon lequel l'individu n'a d'autres choix que de s'adapter pour enfin se conformer aux bouleversements qu'entraînent les technosciences. Ainsi, les Hommes ne font que participer à une logique qui les dépasse. Puisque l'être humain n'a pas le pouvoir d'empêcher la technique de suivre son cours, son seul rôle et responsabilité est de limiter au possible les risques potentiels anticipés et mettre en place des stratégies afin que ces technologies servent au mieux la puissance économique. Il n'est donc pas surprenant que l'analyse de discours ait dévoilé une réflexion exclusivement centrée sur les enjeux de l'IA.

### *Les limites*

En faisant le choix de sélectionner uniquement des discours de vulgarisation, nous avons pu passer à côté de certaines dimensions essentielles à la représentation de l'intelligence des chercheurs en IA au sein de leurs travaux de recherche scientifique. Nous l'avons vu, les stratégies employées dans l'exercice de traduction que constitue la vulgarisation dévoilent au grand jour la représentation sociale de l'émetteur du discours. Cependant, ces mêmes stratégies peuvent tout aussi bien être mobilisées dans le but de dissimuler certains aspects de sa conception du monde. Puisque travaillé, le discours est bien souvent contrôlé par celui qui l'émet. Dans ce contrôle se trouve la possibilité d'altérer la conception de l'intelligence dans un sens plutôt que dans un autre ; pratique bien plus complexe à exécuter au travers des travaux scientifiques. De même, certains raccourcis peuvent être effectués dans une visée de compréhension unanime constituant ainsi une autre limite à notre recherche.

De plus, comme nous l'avons mentionné plus haut, la plupart des chercheurs ne souhaitent pas se soumettre à l'exercice de vulgarisation. Leurs discours ne peuvent donc être pris en compte dans notre analyse délaissant toute représentation qui aurait pu s'avérer divergente. Nous pouvons toutefois émettre l'hypothèse que ces chercheurs sont réfractaires à la divulgation de l'avancée de leurs travaux au grand public justement parce que l'ambition de leur recherche n'est pas d'informer dans une visée démocratique, mais simplement de développer des algorithmes toujours plus performants afin d'élaborer des applications concrètes et optimales. Ce postulat laisse à penser que ces chercheurs se trouveraient en accord parfait, si ce n'est plus, avec la représentation de l'intelligence exposée au terme de ce mémoire.

## Conclusion

« Notre vie se trouve vouée par eux [les décideurs] à l'accroissement de la puissance. Sa légitimation en matière de justice sociale comme de vérité scientifique serait d'optimiser les performances du système, l'efficacité. L'application de ce critère à tous nos jeux ne va pas sans quelques terreurs, douce ou dure : Soyez opératoires, c'est-à-dire commensurables, ou disparaïssez. » (Lyotard, 1979 :8)

Jean-François Lyotard

L'idée de création d'une IA provient à l'origine d'une fascination de notre être en cohérence avec la philosophie humaniste dont les assises sont clairement anthropocentriques. Au fil du développement de l'IA et à travers son histoire, la fascination s'est estompée en curiosité pour ne devenir qu'aliment de la puissance économique. Tout égard envers la recherche de la vérité jusqu'à la considération même de notre propre être se dissipe face à l'attrait au rapport monétaire d'une société néolibérale, ayant pour seule visée la croissance économique. L'approche sociale de l'IA dévoile un nouveau rapport de l'individu à lui-même. L'être humain s'appréhende désormais comme simple machine dotée d'une performance optimale, seule réelle exigence du modèle sociétal dans lequel il évolue. Ainsi, il prend part à un monde toujours plus technique, toujours plus rapidement. Devenu produit bientôt technicisé, l'Homme s'adapte parfaitement aux exigences du modèle sociétal néolibéral en se plaçant lui-même au cœur du marché compétitif. Tout comme l'IA ne sert plus depuis un moment la quête à la découverte de notre nature, l'individu est réduit à sa fonction d'optimisation dont seules ses applications pouvant servir l'intérêt économique sont valorisées. Il participe ainsi à l'extension de ce modèle rendant toute possibilité de rébellion envers celui-ci de plus en plus inaccessible. La puissance de la logique technoscientifique purement opérationnelle nous échappe laissant place à une profonde remise en question des fondements humanistes. La distinction entre l'être humain et de la machine garant de la subjectivité propre à l'être humain nous conférait le pouvoir de changer la société par le biais de la politique et nous garantissait ainsi le contrôle de notre propre condition et l'amélioration sociale de celle-ci. Ces principes fondateurs humanistes n'ont pas leur place au sein du paradigme informationnel, ce qui ultimement impacte la place de l'Homme dans l'univers social ainsi que son rapport à lui-même. Comme l'a parfaitement formulé la sociologue Hannah Arendt, « Un devenir-indifférent à la politique, équivaut à renoncer à la pensée et au jugement, à la lutte pour rendre à nouveau le monde humain. » (Arendt et al., 2016 : 33)

Avec l'arrivée des conséquences du réchauffement climatique et de la surpopulation, le futur semble peu à peu nous échapper. Le manque de temps restant pour gérer cette crise se fait ressentir par une prise de conscience collective inhérente. Il semblerait dorénavant évident qu'il ne suffit plus de laisser à l'évolution naturelle et linéaire du temps le contrôle de notre destinée. En cohérence avec l'idéologie technoscientifique, l'évolution doit alors être exponentielle. On remet ainsi à la technique l'aptitude de nous gouverner. La perception de la technoscience, comme seule capable de nous sauver du naufrage en tant que garant de notre avenir, nous rassure, car nous déresponsabilise. Dans la pensée transhumaniste, la foi en la technique est évidente. Cependant, elle s'étend pour se généraliser à la population par le biais des discours de vulgarisation des autres protagonistes de l'IA. L'idée de trouver la solution dans le problème nous paraît pourtant largement contestable.

Le retour sociohistorique de l'IA, en addition d'une approche constructiviste sur la notion d'intelligence telle qu'elle est assimilée au sein du domaine et donc dans les sociétés néolibérales contemporaines, permet d'entrevoir une autre possibilité pour notre futur. Ce travail de recherche démontre que rien n'est inéluctable puisque tout est construit et que tout reste donc à construire. L'approche choisie pour traiter de ce sujet dévoile que l'évolution de l'IA n'est en rien naturelle, mais dépend au contraire des enjeux sociaux et politiques qui la traversent. Comme le résume le philosophe Cédric Lagandré « Sans recul historique, la fatalité du présent décourage l'action humaine. » (Lagandré, 2009 : 11) L'ambition de ce travail de recherche est de révéler l'illusion d'un avenir fataliste inéluctable soutenue par la croyance en l'autonomie de la technique. Enfin, par-là, peut-être, pourrions-nous aspirer à un autre modèle sociétal ; un modèle dans lequel les Hommes, ensemble, munis des moyens de leur réflexion, de leurs émotions et de leurs instincts peuvent repenser une société plus juste et compatible avec leur humanité profondément sociale, tout en restant en accord avec la planète. Il serait alors concevable d'entrevoir un autre futur possible que celui d'un monde où notre propre individualité est réduite à un simple produit technique doué « d'intelligence », soit selon la représentation que l'on s'en fait en IA, doté d'une capacité de traitement de l'information la plus optimale possible.

## Bibliographie

- ANDRIEU, Bernard, James Russell ANDERSON, LE PARVIS (TARBES) et LES INTELLIGENCES. *Les intelligences animal-homme-machine*, Mont-de-Marsan, Ed. InterUniversitaires, 1992.
- ARCAS, Blaise Agüera y. « Transcription de l'intervention "How computers are learning to be creative" », *Conférence TED*, mai 2016. En ligne au : <[https://www.ted.com/talks/blaise\\_aguera\\_y\\_arcas\\_how\\_computers\\_are\\_learning\\_to\\_be\\_creative/transcript](https://www.ted.com/talks/blaise_aguera_y_arcas_how_computers_are_learning_to_be_creative/transcript)>, consulté le 27 juin 2019.
- ARENDT, Hannah et al. *Qu'est-ce que la politique?*, Paris, Éditions Points, 2016.
- BENBOUZID, Bilel et Dominique CARDON. *Machines prédictives*, 2018.
- BENGIO, Yoshua. « Yoshua Bengio, chercheur d'intelligence artificielle », *La Presse+*, 18 juin 2017a. En ligne au : <[http://plus.lapresse.ca/screens/78ce29ec-a211-444a-b14b-9da2645a0bdf\\_\\_7C\\_\\_0.html](http://plus.lapresse.ca/screens/78ce29ec-a211-444a-b14b-9da2645a0bdf__7C__0.html)>, consulté le 27 juin 2019.
- — —. *Entretien entre Philippe-Audrey Larrue St-Jacques et Yoshua Bengio, roi de l'intelligence artificielle*, 25 avril 2017b. En ligne au : <<https://urbania.ca/article/entretien-entre-philippe-audrey-larrue-st-jacques-et-yoshua-bengio-roi-de-lintelligence-artificielle/>>, consulté le 28 février 2018.
- — —. *L'Intelligence Artificielle avec la sommité mondiale Yoshua Bengio.*, 29 janvier 2017c. En ligne au : <<https://www.youtube.com/watch?v=QbmFisJRQUU>>.
- BENGIO, Yoshua, Aaron COURVILLE et Pascal VINCENT. « Representation Learning: A Review and New Perspectives », *ArXiv12065538 Cs*, 24 juin 2012. En ligne au : <<http://arxiv.org/abs/1206.5538>>, consulté le 17 avril 2019.
- BENSAUDE-VINCENT, Bernadette. « Splendeur et décadence de la vulgarisation scientifique », *Quest. Commun.*, n°17, 30 juin 2010, p. 19-32. <<https://doi.org/10.4000/questionsdecommunication.368>>.
- BERGER, Peter L. et Thomas LUCKMANN. *La construction sociale de la réalité*, Paris, Armand Colin, 1998.
- BINET, Alfred et Théodore SIMON. *L'élaboration du premier test d'intelligence. Oeuvres choisies. II*, Paris, Harmattan, 2004.
- BORING, Edwin G. « Intelligence as the Tests Test It. », dans JENKINS, James J. et Donald G. PATERSON (dir.), *Studies in individual differences: The search for intelligence.*, East Norwalk, Appleton-Century-Crofts, 1961, p. 210-14. <<https://doi.org/10.1037/11491-017>>.
- BOSTROM, Nick. *Nick Bostrom : « Chacun d'entre nous pourra choisir sa propre humanité »*, 1 juillet 2008. En ligne au : <<http://archives.lesechos.fr/archives/2008/Enjeux/00248-066-ENJ.htm>>, consulté le 27 juin 2019.
- — —. *Entretien avec Nick Bostrom, le transhumaniste en chef : article - Revue Argument*, août 2004. En ligne au : <<http://revueargument.ca/article/2005-10-01/331-entretien-avec-nick-bostrom-le-transhumaniste-en-chef.html>>, consulté le 27 juin 2019.
- BOUEE, Charles-Edouard et François ROCHE. *La chute de l'empire humain: mémoires d'un robot*, Paris, Bernard Grasset, 2017.
- BRETON, Philippe. *À l'image de l'Homme: du Golem aux créatures virtuelles*, Paris, Seuil, 1995.
- BRETON, Philippe et Serge PROULX. *L'explosion de la communication: introduction aux théories et aux pratiques de la communication*, Paris, La Découverte, 2016.
- BULLE, Emmanuelle. « Les modèles formels et l'explication en sciences sociales », *Année Sociol.*, vol. 55, n°1, 2005, p. 19-34. <<https://doi.org/10.3917/anso.051.0019>>.
- CARDON, Dominique. *A quoi rêvent les algorithmes: nos vies à l'heure des big data*, La République des idées, 2015.

- CASSAN, Élodie. « La raison chez Descartes, puissance de bien juger », *Le Philosophoire*, vol. 28, n°1, 2007, p. 133-45. <<https://doi.org/10.3917/phoir.028.0133>>.
- DASEN, Pierre R. « The Cross-Cultural Study of Intelligence: Piaget and the Baoule », *Int. J. Psychol.*, vol. 19, n°1-4, janvier 1984, p. 407-34. <<https://doi.org/10.1080/00207598408247539>>.
- DE WOLFF, Kim. « Plastic Naturecultures: Multispecies Ethnography and the Dangers of Separating Living from Nonliving Bodies », *Body Soc.*, vol. 23, n°3, septembre 2017, p. 23-47. <<https://doi.org/10.1177/1357034X17715074>>.
- DECLARATION IA RESPONSABLE. *Déclaration de Montréal pour un développement responsable de l'IA*, Déclaration, Déclaration, Montréal. En ligne au : <<https://www.declarationmontreal-iaresponsable.com/>>, consulté le 27 juin 2019.
- DEEPMIND. « About Us », *DeepMind*. En ligne au : <<https://deepmind.com/about/>>, consulté le 27 juin 2019.
- DELMAS, Corinne. *Sociologie politique de l'expertise*, Paris, La Découverte, 2011.
- DEMANGE, Dominique. « La « définition » aristotélicienne de l'âme », *Le Philosophoire*, vol. n° 21, n°3, 2003, p. 65-85.
- DESCARTES, René et Etienne GILSON. *Discours de la méthode: texte et commentaire*, 6. éd, Paris, Librairie Philosophique Vrin, coll. Bibliothèque des textes philosophiques, 1987.
- DOWEK, Gilles. « Mythologies », dans FRANCE INTER et LIBERATION (dir.), *Intelligence artificielle: enquête sur ces technologies qui changent nos vies.*, Paris, Flammarion, 2018, p. 203-47.
- DREYFUS, Hubert L. « Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian. », *Philos. Psychol.*, vol. 20, n°2, 2007, p. 247-48. — — —. *Alchemy and artificial intelligence*, Rand Corp., 1965.
- DREYFUS, Hubert L., Stuart E. DREYFUS et Tom ATHANASIOU. *Mind over machine: the power of human intuition and expertise in the era of the computer*, New York, Free Press, 1986.
- ESQUIVEL SADA, Daphné. *Le « nanomonde » et le renversement de la distinction entre nature et technique : entre l'artificialisation de la nature et la naturalisation de la technique*, 16 avril 2009. En ligne au : <<https://papyrus.bib.umontreal.ca/xmlui/handle/1866/2755>>, consulté le 22 février 2018.
- FREEMAN, Mallory. « Transcription de l'intervention "Your company's data could help end world hunger" », *Conférence TED*, septembre 2016. En ligne au : <[https://www.ted.com/talks/mallory\\_soldner\\_your\\_company\\_s\\_data\\_could\\_end\\_world\\_hunger/transcript](https://www.ted.com/talks/mallory_soldner_your_company_s_data_could_end_world_hunger/transcript)>, consulté le 27 juin 2019.
- FREITAG, Michel. *L'oubli de la société: pour une théorie critique de la postmodernité*, Laval, Presses de l'Université Laval, coll. Collection Sociologie contemporaine, 2002.
- GANASCIA, Jean-Gabriel. *L'intelligence artificielle: science & techniques*, Paris, Le Cavalier Bleu, coll. Idées reçues, n°138, 2007.
- GARDNER, Howard. *Les intelligences multiples: la théorie qui bouleverse nos idées reçues*, Paris, Retz, 2009.
- GARDNER, Howard, Mindy L. KORNHABER et Warren K. WAKE. *Intelligence: multiple perspectives*, Fort Worth, TX ; Toronto, Harcourt Brace College Publishers, 1996.
- GOOGLE. « AI at Google: our principles », *Google*, 7 juin 2018. En ligne au : <<https://www.blog.google/technology/ai/ai-principles/>>, consulté le 27 juin 2019.
- — —. « AI for Social Good », *Google AI*. En ligne au : <<https://ai.google/social-good/>>, consulté le 27 juin 2019.

- GRUBER, Tom. « Transcription de l'intervention "How AI can enhance our memory, work and social lives" », *Conférence TED*, avril 2017. En ligne au : [https://www.ted.com/talks/tom\\_gruber\\_how\\_ai\\_can\\_enhance\\_our\\_memory\\_work\\_and\\_social\\_lives/transcript](https://www.ted.com/talks/tom_gruber_how_ai_can_enhance_our_memory_work_and_social_lives/transcript), consulté le 27 juin 2019.
- HARAWAY, Donna Jeanne, Laurence ALLARD, Delphine GARDEY et Nathalie MAGNAN. *Manifeste cyborg et autres essais: sciences, fictions, féminismes*, Paris, Exils, coll. Essais, 2007.
- HARRIS, Sam. « Transcription de l'intervention "Can we build AI without losing control over it?" », *Conférence TED*, juin 2016. En ligne au : [https://www.ted.com/talks/sam\\_harris\\_can\\_we\\_build\\_ai\\_without\\_losing\\_control\\_over\\_it?referrer=playlist-talks\\_on\\_artificial\\_intelligen#t-849203](https://www.ted.com/talks/sam_harris_can_we_build_ai_without_losing_control_over_it?referrer=playlist-talks_on_artificial_intelligen#t-849203), consulté le 27 juin 2019.
- HEBB, Donald O. *The organization of behavior: a neuropsychological theory*, 11. [print.], New York, Wiley, 1974.
- HOWARD, Jeremy. « Transcription de l'intervention : "The wonderful and terrifying implications of computers that can learn" », *Conférence TED*, décembre 2014. En ligne au : [https://www.ted.com/talks/jeremy\\_howard\\_the\\_wonderful\\_and\\_terrifying\\_implications\\_of\\_computers\\_that\\_can\\_learn](https://www.ted.com/talks/jeremy_howard_the_wonderful_and_terrifying_implications_of_computers_that_can_learn), consulté le 28 juin 2019.
- HUMANITY+. « Transhumanist FAQ », *Humanity+*. En ligne au : <https://humanityplus.org/philosophy/transhumanist-faq/>, consulté le 28 juin 2019.
- ICTC. *Artificial Intelligence in Canada : Where do we stand?*, Information and communication technology concil, 2015.
- JODELET, Denise. « Représentations sociales : un domaine en expansion », dans *Les représentations sociales*, 7e éd., Presses Universitaires de France, 2003, p. 45-78. En ligne au : <https://www.cairn.info/les-representations-sociales--9782130537656-page-45.htm>, consulté le 28 juin 2019.
- JOSSET, Raphaël. « La destinée cybernétique du monde », *Sociétés*, n°131, 4 octobre 2016, p. 9-17. <https://doi.org/10.3917/soc.131.0009>.
- JULIA, Luc. *Nouveau monde. Luc Julia : « L'intelligence artificielle n'existe pas et la voiture autonome n'existera jamais »*, 25 janvier 2019. En ligne au : [https://www.francetvinfo.fr/replay-radio/nouveau-monde/nouveau-monde-luc-julia-lintelligence-artificielle-nexiste-pas-et-la-voiture-autonome-nexistera-jamais\\_3140767.html](https://www.francetvinfo.fr/replay-radio/nouveau-monde/nouveau-monde-luc-julia-lintelligence-artificielle-nexiste-pas-et-la-voiture-autonome-nexistera-jamais_3140767.html), consulté le 27 juin 2019.
- KURZWEIL, Ray. « Human Body Version 2.0 », *Kurzweil Accel. Intell.*, 13 février 2003. En ligne au : <http://www.kurzweilai.net/human-body-version-20>, consulté le 28 février 2018.
- LAFONTAINE, Céline. « The Cybernetic Matrix of 'French Theory' », *Theory Cult. Soc.*, vol. 24, n°5, 1 septembre 2007, p. 27-46. <https://doi.org/10.1177/0263276407084637>.
- — —. *L'empire cybernétique: des machines à penser à la pensée machine: essai*, Paris, Seuil, 2004.
- LAGANDRE, Cédric. *La société intégrale*, Paris, Climats, 2009.
- LATOUR, Bruno. *Nous n'avons jamais été modernes: essai d'anthropologie symétrique*, Nachdr., Paris, Editions La Découverte [u.a.], 2010.
- LE DEVEDEC, Nicolas. *La société de l'amélioration: la perfectibilité humaine des lumières au transhumanisme*, Montréal, Liber, 2015a.
- — —. « Retour vers le futur transhumaniste », *Esprit*, vol. Novembr, n°11, 2015b, p. 89. <https://doi.org/10.3917/espri.1511.0089>.

- LE FIGARO. « Cinq idées reçues sur l'intelligence artificielle », *Le Figaro*, 19 janvier 2018. En ligne au : <<http://www.lefigaro.fr/secteur/high-tech/2018/01/19/32001-20180119ARTFIG00276-cinq-idees-recues-sur-l-intelligence-artificielle.php>>, consulté le 28 juin 2019.
- LE MONDE. « En quoi l'intelligence artificielle favorise-t-elle l'emploi ? », *Le Monde*, 29 août 2018. En ligne au : <[https://www.lemonde.fr/emploi/article/2018/08/29/en-quoi-l-intelligence-artificielle-favorise-t-elle-l-emploi\\_5347616\\_1698637.html](https://www.lemonde.fr/emploi/article/2018/08/29/en-quoi-l-intelligence-artificielle-favorise-t-elle-l-emploi_5347616_1698637.html)>, consulté le 27 juin 2019.
- LECUN, Yann. « Qu'est ce que l'IA? », dans *Intelligence artificielle: enquête sur ces technologies qui changent nos vies.*, Paris, Flammarion, 2018, p. 13-49.
- — —. *Les enjeux de la recherche en Intelligence Artificielle.*, Chaire informatique et sciences numérique, Collège de France, Le «Deep learning», une révolution en Intelligence artificielle, 4 février 2016. En ligne au : <[https://www.college-de-france.fr/media/presse/UPL7132967935621487999\\_Dossier\\_YLeCun.pdf](https://www.college-de-france.fr/media/presse/UPL7132967935621487999_Dossier_YLeCun.pdf)>.
- LEROI-GOURHAN, André. *La mémoire et les rythmes*, Repr, Paris, Michel, coll. Le geste et la parole, n°2, 1998.
- LYOTARD, Jean-François. *La condition postmoderne: rapport sur le savoir*, Paris, Éditions de Minuit, coll. Collection Critique, 1979.
- MCCARTHY, John. « Epistemological Problems of Artificial Intelligence », dans *Readings in Artificial Intelligence*, Elsevier, 1981, p. 459-65. <<https://doi.org/10.1016/B978-0-934613-03-3.50035-0>>.
- MCCARTHY, John et Patrick J. HAYES. « Some Philosophical Problems from the Standpoint of Artificial Intelligence », dans *Machine Intelligence 4*, Edinburgh University Press, 1969, p. 463-502.
- MCCARTHY, John, Marvin L. MINSKY, Nathaniel ROCHESTER et Claude E. SHANNON. « A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955 », *AI Mag.*, vol. 27, n°4, 15 décembre 2006, p. 12. <<https://doi.org/10.1609/aimag.v27i4.1904>>.
- MCCORDUCK, Pamela. *Machines who think: a personal inquiry into the history and prospects of artificial intelligence*, 25th anniversary update, Natick, Mass, A.K. Peters, 2004.
- MCCULLOCH, Warren et Walter PITTS. « A Logical Calculus of Ideas Immanent in Nervous Activity », *Bull. Math. Biophys.*, n°5, 1943, p. 115-33.
- MERMILLOD, Martial. *Réseaux de neurones biologiques et artificiels: vers l'émergence de systèmes artificiels conscients?*, 2016.
- MERVEILLE, Nicolas. *IA Montréal : ville intelligente, ville intelligible*, Conférence CERIU, Université de Montréal, Enjeux politiques de l'Intelligence Artificielle, 2018.
- MILA. « IA pour l'humanité », *MILA*. En ligne au : <<https://mila.quebec/ia-dans-la-societe/>>, consulté le 28 juin 2019.
- MINSKY, Marvin. « Transcription de l'intervention "Health and the human mind" », *Conférence TED*, février 2003. En ligne au : <[https://www.ted.com/talks/marvin\\_minsky\\_on\\_health\\_and\\_the\\_human\\_mind/transcript](https://www.ted.com/talks/marvin_minsky_on_health_and_the_human_mind/transcript)>, consulté le 27 juin 2019.
- — —. *La société de l'esprit*, Paris, InterEditions, 1988.
- MINSKY, Marvin et Seymour A. PAPER. *Perceptrons: an introduction to computational geometry*, 2. print. with corr, Cambridge/Mass., The MIT Press, 1972.
- MITCHELL, Margaret. « Transcription de l'intervention "How we can build AI to help humans, not hurt us" », *Conférence TED*, octobre 2017. En ligne au : <[https://www.ted.com/talks/margaret\\_mitchell\\_how\\_we\\_can\\_build\\_ai\\_to\\_help\\_humans\\_not\\_hurt\\_us/transcript](https://www.ted.com/talks/margaret_mitchell_how_we_can_build_ai_to_help_humans_not_hurt_us/transcript)>, consulté le 27 juin 2019.

- MUGNY, Gabriel et Felice CARUGATI. *Social representations of intelligence*, Cambridge, Cambridge University Press, 2009.
- MUSK, Elon. « Neuralink: Elon Musk’s Elusive Brain-Computer Firm Just Made a Big Reveal », *Inverse*. En ligne au : <<https://www.inverse.com/article/57607-neuralink-elon-musk-s-elusive-brain-computer-firm-just-made-a-big-reveal>>, consulté le 1 août 2019.
- NEUMANN, John von. *First Draft of a Report on the EDVAC*, Between the United States Army Ordnance Department and the University of Pennsylvania, Contract n° W-670-ORD-4926, Moore School of Electrical Engineering University of Pennsylvania, 1945.
- NEWELL, Allen et Herbert A. SIMON. *Human problem solving*, Englewood Cliffs, N.J., Prentice-Hall, 1972.
- NICOLAS, Serge et Ludovic FERRAND. *Histoire de la psychologie scientifique*, Bruxelles, De Boeck, 2008.
- NILSSON, Nils J. *The quest for artificial intelligence: a history of ideas and achievements*, Cambridge ; New York, Cambridge University Press, 2010.
- NSTC. *Preparing for the future of Intelligence*, National Science and Technology Council & Office of Science and Technology Policy. Executive office of the president of the United States, 2016.
- OVIDE. « De Pygmalion épris d’une statue d’ivoire dont Aphrodite fit pour lui une femme. », dans *Métamorphoses*, Flammarion, liv. X, coll. GF, Paris, 1966, p. versets 243-297.
- PERLER, Dominik. « La théorie cartésienne du jugement. Remarques sur la IV<sup>e</sup> méditation », *Études Philos.*, vol. 71, n°4, 2004, p. 461-83. <<https://doi.org/10.3917/leph.044.0461>>.
- PIAGET, Jean. *La psychologie de l’intelligence*, Paris, A. Colin, 1988.
- PICQ, Pascal. « Homo et la fin des certitudes », *Communications*, vol. 95, n°2, 2014, p. 19. <<https://doi.org/10.3917/commu.095.0019>>.
- PINEAU, Joëlle. « Les machines peuvent être créatives », assure Joëlle Pineau, à la tête du laboratoire FAIR de Facebook à Montréal, 29 juin 2018. En ligne au : <<https://www.france24.com/fr/>>, consulté le 27 juin 2019.
- POMIAN-SAADJIAN, Joanna. « Aux origines de l’Intelligence Artificielle : H. A. Simon en père fondateur », *Quaderni*, vol. 1, n°1, 1987, p. 9-25. <<https://doi.org/10.3406/quad.1987.2093>>.
- PROUST, Joëlle. « Redéfinir l’humain. Pour une convergence des sciences de l’homme. », *Le Débat*, vol. 180, n°3, 2014, p. 56-69. <<https://doi.org/10.3917/deba.180.0056>>.
- REDMON, Joseph. « Transcription de l’intervention : “How computers learn to recognize objects instantly” », avril 2017. En ligne au : <[https://www.ted.com/talks/joseph\\_redmon\\_how\\_a\\_computer\\_learns\\_to\\_recognize\\_objects\\_instantly](https://www.ted.com/talks/joseph_redmon_how_a_computer_learns_to_recognize_objects_instantly)>, consulté le 28 juin 2019.
- ROBILLARD, Jean. « Qu’y a-t-il d’intelligent en intelligence artificielle? », *Manuscr. En Circ. Libre Paginé*, 2019, p. 21.
- ROSZAK, Theodore. *The cult of information: the folklore of computers and the true art of thinking*, 1st ed, New York, Pantheon, 1986.
- SEARLE, John R. « Minds, brains, and programs », *Behav. Brain Sci.*, vol. 3, n°3, septembre 1980, p. 417-24. <<https://doi.org/10.1017/S0140525X00005756>>.
- SHANNON, C. E. « A Mathematical Theory of Communication », *SIGMOBILE Mob Comput Commun Rev*, vol. 5, n°1, janvier 2001, p. 3–55. <<https://doi.org/10.1145/584091.584093>>.
- SIMON, Herbert Alexander. *The sciences of the artificial*, 3. ed., [Nachdr.], Cambridge, Mass., MIT Press, 2008.
- — —. *The shape of automation for men and management*, Harper & Row, 1965.

- TEGMARK, Max. « Transcription de l'intervention "How to get empowered, not overpowered, by AI" », *Conférence TED*, avril 2018. En ligne au : [https://www.ted.com/talks/max\\_tegmark\\_how\\_to\\_get\\_empowered\\_not\\_overpowered\\_by\\_ai](https://www.ted.com/talks/max_tegmark_how_to_get_empowered_not_overpowered_by_ai), consulté le 27 juin 2019.
- TESTART, Jacques. « *L'abeille est plus intelligente qu'un ordinateur* », 26 juin 2018. En ligne au : <http://jacques.testart.free.fr/public/pdf/texte1040.pdf>.
- THURSTONE, L. L. *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.*, Chicago, University of Chicago Press, 1935. <https://doi.org/10.1037/10018-000>.
- TUAL, Morgane. « Intelligence artificielle : une machine peut-elle ressentir de l'émotion ? », 12 octobre 2015. En ligne au : [https://www.lemonde.fr/pixels/article/2015/10/12/intelligence-artificielle-une-machine-peut-elle-ressentir-de-l-emotion\\_4787837\\_4408996.html](https://www.lemonde.fr/pixels/article/2015/10/12/intelligence-artificielle-une-machine-peut-elle-ressentir-de-l-emotion_4787837_4408996.html), consulté le 4 décembre 2018.
- TURING, Alan. « Computing Machinery and Intelligence », *Mind*, vol. LIX, n°236, 1950, p. 433-60. <https://doi.org/10.1093/mind/LIX.236.433>.
- . « On Computable Numbers, with an Application to the Entscheidungsproblem », *Proc. Lond. Math. Soc.*, vol. 43, 1937, p. 230-65.
- VILLANI, Cédric. *Mission Villani sur l'Intelligence Artificielle : Qu'est ce que l'Intelligence Artificielle*, Livret de vulgarisation, Livret de vulgarisation, 2018.
- WATSON, John. « Psychology as the Behaviorist views it », *Psychol. Rev.*, n°20, 1913, p. 158-77.
- WEINBERG, Gill. « Transcription de l'intervention "Can robots be creative?" », *Conférence TEDed*, mars 2015. En ligne au : [https://www.ted.com/talks/gil\\_weinberg\\_can\\_robots\\_be\\_creative/transcript](https://www.ted.com/talks/gil_weinberg_can_robots_be_creative/transcript), consulté le 27 juin 2019.
- WIENER, Norbert. *Cybernétique et société: l'usage humain des êtres humains*---, Paris, Union générale d'éditions, coll. Monde en 10/18 ; 56, 1962.
- WIENER, Norbert, Ronan LE ROUX, Robert VALLEE et Nicole VALLEE. *La cybernétique information et régulation dans le vivant et la machine*, Paris, Éd. du Seuil, 2014.
- WIENER, Norbert, Christophe WALL-ROMANA, Charles MOPSIK et Gershom Gerhard SCHOLEM. *God & [et] golem inc.: sur quelques points de collision entre cybernétique et religion*, Nîmes, Ed. de l'Eclat, 2001.
- WISSNER-GROSS, Alex. « Transcription de l'intervention "A new equation for intelligence" », *Conférence TED*, novembre 2013. En ligne au : [https://www.ted.com/talks/alex\\_wissner\\_gross\\_a\\_new\\_equation\\_for\\_intelligence/transcript](https://www.ted.com/talks/alex_wissner_gross_a_new_equation_for_intelligence/transcript), consulté le 27 juin 2019.
- X, THE MOONSHOT FACTORY. « X - Brain », *X Moonshot Fact.* En ligne au : <https://x.company>, consulté le 27 juin 2019.