

*Pierre Lévy\**

## Vers un changement de paradigme en intelligence artificielle

### Introduction

Examinons d'abord comment le terme "intelligence artificielle"<sup>1</sup> (IA) est utilisé dans la société en général, par exemple par les journalistes et les publicitaires. L'observation historique montre que l'on a tendance à classer dans l'intelligence artificielle les applications considérées comme "avancées" à l'époque où elles apparaissent. Mais quelques années plus tard ces mêmes applications seront le plus souvent réinterprétées comme appartenant à l'informatique ordinaire. Par exemple, la reconnaissance optique de caractères, perçue comme de l'IA à l'origine, est aujourd'hui considérée comme normale et silencieusement intégrée dans de nombreux logiciels. Une machine capable de jouer aux échecs était célébrée comme un exploit technique jusqu'aux années 1970, mais l'on peut aujourd'hui télécharger un programme d'échecs gratuit sur son smartphone sans que nul ne s'en étonne. De plus, selon que l'IA est en vogue (comme aujourd'hui) ou déconsidérée (comme dans les années 1990-2000), les efforts de marketing mettront ce terme en avant ou le remplaceront par d'autres. Par exemple, les "systèmes experts" des années 1980 deviennent les anodines "règles d'affaire" des années 2000. C'est ainsi que des techniques ou des concepts identiques changent de dénomination selon les modes, rendant la perception du domaine et de son évolution particulièrement opaque.

Quittons maintenant le vocabulaire du journalisme ou du marketing pour nous intéresser à la discipline académique. L'intelligence artificielle désigne depuis les années 1950 la branche de l'informatique qui se préoccupe de modéliser et de simuler l'intelligence humaine dans

\* Université de Montréal. CEO de INLEKT Metadata Inc

<sup>1</sup> L'expression "Intelligence artificielle" fut utilisée pour la première fois en 1956, lors d'une conférence du Dartmouth College à Hanover, New Hampshire. Participaient notamment à cette conférence l'informaticien et chercheur en sciences cognitives Marvin Minsky (Turing Award 1969) et l'inventeur du langage de programmation LISP John McCarthy (Turing Award 1971).

son ensemble plutôt que de résoudre tel ou tel problème particulier. La modélisation informatique de l'intelligence humaine est un but scientifique légitime qui a eu et continuera à avoir des retombées théoriques et pratiques considérables. Néanmoins, échaudés par les prévisions enthousiastes, mais démenties par les faits, des débuts de la discipline, la plupart des chercheurs du domaine ne croient pas que l'on construira bientôt des machines intelligentes *autonomes*. Beaucoup de recherches dans ce domaine – ainsi que la plupart des applications pratiques – visent d'ailleurs une *augmentation de la cognition humaine*<sup>2</sup> plutôt que sa reproduction mécanique. Par opposition au programme de recherche orienté vers la construction d'une intelligence artificielle générale autonome, j'ai défendu dans mon livre *La Sphère Sémantique* l'idée d'une intelligence artificielle au service de l'intelligence collective et du développement humain. Je poursuis ici cette ligne de pensée.

D'un point de vue technique, l'IA se partage en deux grandes branches: symbolique et statistique. Un algorithme d'IA statistique "apprend" à partir des données qu'on lui fournit. Il simule donc (imparfaitement, nous le verrons plus bas), la dimension *inductive* du raisonnement humain. Par contraste, l'IA symbolique n'apprend pas à partir des données, mais dépend de la formalisation logique de la connaissance d'un domaine par des ingénieurs. Comparée à l'IA statistique, elle demande donc en principe une quantité plus importante de travail intellectuel humain. Un algorithme d'IA symbolique applique aux données les règles qu'on lui a données. Il simule donc plutôt la dimension *déductive* du raisonnement humain. Je vais successivement passer en revue ces deux grandes branches de l'IA, en m'attachant plus particulièrement à souligner leurs limites.

## L'IA statistique et ses limites

### L'IA neuronale

La branche *statistique* de l'IA entraîne des algorithmes à partir d'énormes masses de données pour les rendre capable de reconnaître des formes visuelles, sonores, linguistiques ou autres. C'est ce que l'on appelle l'apprentissage automatique ou *machine learning*. Lorsque l'on parle d'IA

<sup>2</sup> *L'augmentation cognitive* (plutôt que *l'imitation* de l'intelligence humaine) était l'orientation principale de nombre des pionniers de l'informatique et du Web. Voir par exemple: V. Bush, *As We May Think*, Atlantic Monthly, Boston 1945; J. Licklider, *Man-Computer Symbiosis*, in "IRE Transactions on Human Factors in Electronics", 1, 1960, pp. 4-11; D. Engelbart, *Augmenting Human Intellect, Technical Report*, Stanford Research Institute, Stanford 1962; T. Berners-Lee, *Weaving the Web*, Harper, San Francisco 1999.

en 2021, c'est généralement pour désigner ce type de technique. On l'a vu, l'IA statistique économise le travail humain si on la compare à l'IA symbolique. Il suffit de fournir à un algorithme d'apprentissage automatique un ensemble de données d'entraînement pour qu'un programme de reconnaissance de formes s'écrive tout seul. Si l'on donne par exemple à une IA statistique des millions d'images de canards accompagnées d'étiquettes précisant que l'image représente un canard, elle apprend à reconnaître un canard et, à l'issue de son entraînement, elle sera capable de coller elle-même l'étiquette "canard" sur une image non catégorisée de ce volatile. Personne n'a expliqué à la machine *comment* reconnaître un canard: on s'est contenté de lui fournir des exemples. La traduction automatique répond au même principe: on donne à une IA statistique des millions de textes dans une langue A accompagnés de leur traduction dans une langue B. Entraîné sur ces exemples, le système apprend à traduire un texte de la langue A dans la langue B. C'est ainsi que fonctionnent des algorithmes de traduction automatique comme DeepL ou Google Translate. Pour prendre un exemple dans un autre domaine, l'IA statistique utilisée pour conduire les "véhicules autonomes" fonctionne également en *appariant deux ensembles de données*: des images de la route sont mises en correspondance avec des *actions* telles qu'accélérer, freiner, tourner, etc. En somme, l'IA statistique établit une correspondance (*mapping*) entre un ensemble de données et un ensemble d'étiquettes (cas de la reconnaissance de forme) ou bien entre deux ensembles de données (cas de la traduction ou des véhicules autonomes). Elle excelle donc dans les tâches de catégorisation, de reconnaissance de forme et d'appariement réflexe entre données perceptives et données motrices.

Dans sa version la plus perfectionnée, l'IA statistique repose sur des modèles de *réseaux neuronaux* qui simulent grossièrement le mode d'apprentissage du cerveau. On parle d'apprentissage "profond" (*deep learning* en anglais) pour qualifier ces modèles parce qu'ils reposent sur *plusieurs couches* superposées de neurones formels. Les réseaux neuronaux représentent le sous-domaine le plus complexe et le plus avancé de l'IA statistique. L'intelligence artificielle de type neuronal existe depuis l'origine de l'informatique, comme l'illustrent les recherches de McCulloch<sup>3</sup> dans les années 1940 et 50, de Franck Rosenblatt<sup>4</sup> et Marvin Minsky<sup>5</sup> dans les

<sup>3</sup> W. McCulloch, W. Pitts, *A Logical Calculus of Ideas Immanent in Nervous Activity*, in "Bulletin of Mathematical Biophysics", 5, 1943, pp. 115-133.

<sup>4</sup> Frank Rosenblatt est l'inventeur du Perceptron, qui peut être considéré comme le premier système d'apprentissage automatique basé sur un réseau neuro-mimétique.

<sup>5</sup> Le mémoire de doctorat de 1954 de Marvin Minsky était intitulé: "*Theory of neural-analog reinforcement systems and its application to the brain-model problem*". Minsky critiquera le perceptron de Frank Rosenblatt dans son livre: M. Minsky, *Perceptron*, MIT Press, Cambridge MA 1969, écrit avec Seymour Papert et poursuivra par la suite le pro-

années 1950 et de von Foerster<sup>6</sup> dans les années 1960 et 70. D'importants travaux dans ce domaine ont également été menés dans les années 1980, impliquant notamment David Rumelhart<sup>7</sup> et Geoffrey Hinton<sup>8</sup>, mais toutes ces recherches ont eu peu de succès pratique avant les années 2010.

Outre certains perfectionnements scientifiques des modèles, deux facteurs indépendants des progrès de la théorie expliquent que les réseaux neuronaux soient de plus en plus utilisés: la disponibilité d'énormes masses de données et l'augmentation de la puissance de calcul. À partir de la seconde décennie du XXI<sup>e</sup> siècle, les organisations s'engagent dans la transformation numérique et une part croissante de la population mondiale utilise le Web. Tout cela génère de gigantesques flux de données. Les informations ainsi produites sont traitées par les grandes plateformes numériques dans des *centres de données* (le "*cloud*") qui concentrent une puissance de calcul inouïe. Au début du XXI<sup>e</sup> siècle, les réseaux neuronaux étaient implémentés par des processeurs conçus à l'origine pour le calcul graphique, mais les centres de données des grandes plateformes utilisent maintenant des processeurs spécialement destinés à l'apprentissage neuronal. C'est ainsi que des modèles théoriques intéressants, mais peu pratiques, du XX<sup>e</sup> siècle sont soudain devenus pertinents au XXI<sup>e</sup> siècle au point de soutenir une nouvelle industrie.

### Des rendements décroissants

Néanmoins, après les avancées foudroyantes des années 2010 en matière d'apprentissage automatique par les réseaux neuronaux, les progrès semblent marquer le pas depuis quelques années. En effet, pour obtenir des performances marginalement meilleures, il faut désormais multiplier par plusieurs ordres de grandeur la taille des ensembles de données et la puissance de calcul utilisée pour entraîner les modèles. Nous avons déjà atteint l'époque des *rendements cognitifs décroissants* pour l'IA neuronale. Il est donc temps de s'interroger sur les limites de cet ensemble de techniques et d'envisager sérieusement un changement de paradigme.

Les principaux problèmes portent sur la qualité des données d'entraî-

gramme de recherche de l'IA symbolique; M. Minsky, *The Society of Mind*, Simon and Schuster, New York 1986, résume bien son approche de la cognition humaine comme une émergence à partir de l'interaction de multiples modules cognitifs aux fonctions variées.

<sup>6</sup> Heinz Von Foerster était directeur du Biological Computer Laboratory entre 1958 et 1975 à l'Université de l'Illinois Urbana-Champaign et il fut le secrétaire des conférences Macy où s'élabora la cybernétique (1941-1960).

<sup>7</sup> D. Rumelhart, G. Hinton, E. Geoffrey, R. Williams, *Learning representations by back-propagating errors*, in "Nature", 323, 1985, pp. 533-536.

<sup>8</sup> G. Hinton en 1986 a été reconnu pour ses travaux pionniers par un prix Turing obtenu avec Yann LeCun et Yoshua Bengio en 2018.

nement, l'absence de modélisation causale, le caractère inexplicable des résultats, l'absence de généralisation, la cécité par rapport au sens des données et les difficultés d'accumulation et d'intégration des connaissances.

### **La qualité des données d'entraînement**

Un ingénieur de Google aurait déclaré plaisamment: "Chaque fois que nous licencions un linguiste, notre performance en traduction automatique s'améliore". Mais bien que l'IA statistique soit réputée peu gourmande en travail humain, les risques de biais et d'erreurs soulignés par des utilisateurs de plus en plus sourcilleux poussent à mieux sélectionner les données d'entraînement et à les étiqueter d'une manière plus soigneuse. Or cela demande du temps et de l'expertise humaine, bien qu'il s'agisse précisément des facteurs que l'on espérait éliminer.

### **L'absence d'hypothèses causales explicites**

Tous les cours de statistiques commencent par une mise en garde contre la confusion entre corrélation et causalité. Une corrélation entre A et B ne prouve pas que A est la cause de B. Il peut s'agir d'une coïncidence, ou bien B peut être la cause de A, ou bien un facteur C non pris en compte par le recueil de données est la véritable cause de A et B, sans parler de toutes les relations systémiques complexes imaginables impliquant A et B. Or l'apprentissage automatique repose sur des appariements de données, c'est-à-dire sur des corrélations. La notion de causalité est étrangère à l'IA statistique, comme à de nombreuses techniques d'analyse de données massives, bien que des hypothèses causales interviennent souvent de manière implicite dans le choix des ensembles de données et de leur catégorisation. En somme, l'IA neuronale contemporaine n'est pas capable de distinguer les causes des effets<sup>9</sup>. Pourtant, quand on utilise l'IA pour l'aide à la décision et plus généralement pour s'orienter dans des domaines pratiques, il est indispensable de posséder des modèles causaux explicites, car les actions efficaces doivent bel et bien intervenir *sur les causes*. Dans une démarche scientifique intégrale, les mesures statistiques et les hypothèses causales s'inspirent et se contrôlent réciproquement. Ne considérer que les corrélations statistiques relève d'une dangereuse hémiplegie cognitive. Quant à la pratique répandue qui consiste à garder ses théories causales implicites, elle interdit de les relativiser, de les

<sup>9</sup> J. Pearl, D. Mackenzie, *The Book of Why. The new science of cause and effect*, Basic Books, New York 2019.

comparer avec d'autres théories, de les généraliser, de les partager, de les critiquer et de les perfectionner<sup>10</sup>.

### Des résultats inexplicables

Le fonctionnement des réseaux neuronaux est opaque. Des millions d'opérations transforment de manière incrémentale la force des connexions dans des assemblées de neurones comportant des centaines de couches. Comme leurs résultats ne peuvent être expliqués ni justifiés de manière conceptuelle, c'est-à-dire sur un mode compréhensible par des humains, il est difficile de faire confiance à ces modèles. Cette absence d'explication devient inquiétante lorsque les machines prennent des décisions financières, judiciaires, médicales ou liées à la conduite de véhicules autonomes, sans parler des applications militaires. Pour surmonter cet obstacle, et parallèlement au développement de l'éthique de l'intelligence artificielle, de plus en plus de chercheurs explorent le nouveau champ de recherche de "l'IA explicable" (*explainable AI*).

### L'absence de généralisation

L'IA statistique se présente à première vue comme une forme de raisonnement inductif, c'est-à-dire comme une capacité à inférer des règles générales à partir d'une multitude de cas particuliers. Pourtant, les systèmes d'apprentissage automatique contemporains ne parviennent pas à généraliser au-delà des limites des données d'entraînement qui leur ont été fournies. Non seulement nous – les humains – sommes capables de généraliser à partir de quelques exemples, alors qu'il faut des millions de cas pour entraîner des machines, mais nous pouvons abstraire et conceptualiser ce que nous avons appris tandis que l'apprentissage automatique ne parvient pas à extrapoler et encore moins à conceptualiser. Il reste au niveau d'un apprentissage purement réflexe, étroitement circonscrit par l'espace des exemples qui l'ont alimenté.

### La cécité au sens

Alors que les performances en traduction ou en écriture automatique (tel qu'illustré par le programme GPT-3<sup>11</sup>) progressent, les machines ne

<sup>10</sup> Outre les travaux de Judea Pearl sur l'importance de la modélisation causale en IA, rappelons les thèses du philosophe Karl Popper sur les limites du raisonnement inductif et des statistiques: K. Popper, *Objective Knowledge: An Evolutionary Approach*, Clarendon Press, Oxford 1972.

<sup>11</sup> OpenAI: <https://openai.com/blog/gpt-3-apps/>.

parviennent toujours pas à comprendre le *sens* des textes qu'elles traduisent ou rédigent. Leurs réseaux neuronaux ressemblent au cerveau d'un perroquet mécanique capable d'imiter des performances linguistiques sans avoir la moindre idée du contenu des textes. La succession des mots dans une langue ou leur correspondance d'une langue à l'autre sont bien maîtrisées, mais les textes "reconnus" n'alimentent pas de représentations utilisables des situations ou des domaines de connaissance dont ils traitent.

### **Les difficultés d'accumulation et d'intégration des connaissances par l'IA statistique**

Privée de concepts, l'IA statistique parvient difficilement à *accumuler* des connaissances. A fortiori, *l'intégration* de savoirs de divers champs d'expertise semble hors de portée. Cette situation ne favorise pas les échanges de connaissances entre machines. Il faut donc souvent recommencer à zéro pour chaque nouveau projet. Signalons néanmoins l'existence de modèles de traitement des langues naturelles comme BERT qui sont pré-entraînés sur des données générales et qu'il est ensuite possible de spécialiser dans des domaines particuliers. Une forme limitée de capitalisation est donc atteignable. Mais il reste impossible d'intégrer directement à un système neuro-mimétique l'ensemble des connaissances objectives accumulées par l'humanité depuis quelques siècles<sup>12</sup>.

### **L'IA symbolique et ses limites**

La branche *symbolique* de l'IA correspond à ce qui a été successivement appelé dans les soixante-dix dernières années: réseaux sémantiques, systèmes à base de règles, bases de connaissances, systèmes experts, web sémantique et, plus récemment, graphes de connaissance. Depuis ses origines dans les années 1940-50, une bonne partie de l'informatique appartient de fait à l'IA symbolique.

L'IA symbolique code la connaissance humaine de manière explicite sous forme de réseaux de relations entre catégories et de règles logiques donnant prise au raisonnement automatique. Ses résultats sont donc plus facilement explicables que ceux de l'IA statistique.

<sup>12</sup> Voir le récent rapport du Center for Research on Foundation Models (CRFM) at the Stanford Institute for Human-Centered Artificial Intelligence (HAI), intitulé *On the Opportunities and Risks of Foundation Models* et qui commence ainsi: "AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) that are trained on broad data at scale and are adaptable to a wide range of downstream tasks".

## Les difficultés d'accumulation et d'intégration des connaissances par l'IA symbolique

L'IA symbolique fonctionne bien dans les micromondes fermés des jeux ou des laboratoires, mais se trouve rapidement dépassée dans les environnements ouverts qui ne répondent pas à un petit nombre de règles strictes. La plupart des programmes d'IA symbolique utilisés dans des environnements de travail réels ne résolvent de problèmes que dans un domaine étroitement limité, qu'il s'agisse de diagnostic médical, de dépannage de machines, de conseil en investissement ou autre. Un "système expert" fonctionne de fait comme un *médium* d'encapsulation et de distribution d'un savoir-faire particulier, qui peut être distribué partout où on en a besoin. La compétence pratique devient alors disponible même en l'absence de l'expert humain.

À la fin des années 1980, à la suite d'une série de promesses inconsidérées suivies de déceptions, commence ce que l'on a appelé "l'hiver" de l'intelligence artificielle (toutes tendances confondues). Pourtant, les mêmes procédés continuent à être utilisés pour résoudre le même type de problèmes. On a seulement renoncé au programme de recherche général dans lequel ces méthodes s'inscrivaient. C'est ainsi qu'au début du XXI<sup>e</sup> siècle, les *règles d'affaires* des logiciels d'entreprise et les *ontologies* du Web Sémantique ont succédé aux *systèmes experts* des années 1980. Malgré les changements de nom, il est aisé de reconnaître dans ces nouvelles spécialités les procédés de la bonne vieille IA symbolique.

À partir du début du XXI<sup>e</sup> siècle, le "Web sémantique"<sup>13</sup> s'est donné pour finalité d'exploiter les informations disponibles dans l'espace ouvert du Web. Afin de rendre les données lisibles par les ordinateurs, on organise les différents domaines de connaissance ou de pratique en modèles cohérents. Ce sont les "ontologies", qui ne peuvent que reproduire le *cloisonnement* logique des décennies précédentes, malgré le fait que les ordinateurs soient maintenant beaucoup plus interconnectés.

Malheureusement, nous retrouvons dans l'IA symbolique les mêmes difficultés d'intégration et d'accumulation des connaissances que dans l'IA statistique<sup>14</sup>. Ce cloisonnement entre en opposition avec le projet originel de l'intelligence artificielle comme discipline scientifique, qui veut modéliser l'intelligence humaine en général et qui tend normalement vers une *accumulation* et une *intégration* des connaissances mobilisables par les machines.

<sup>13</sup> Voir le site: <https://www.w3.org/standards/semanticweb/#>.

<sup>14</sup> L'intégration des connaissances existantes dans les systèmes d'IA est un des principaux objectifs du "Wolfram Language" de Stephen Wolfram. Voir: <https://www.wolfram.com/language/principles/>.

Malgré le cloisonnement de ses modèles, l'IA symbolique est cependant un peu mieux lotie que l'IA statistique en matière d'accumulation et d'échange. Un nombre croissant d'entreprises, à commencer par les grandes compagnies du Web, organisent leurs bases de données au moyen d'un graphe de connaissance constamment amélioré et augmenté. Par ailleurs, Wikidata<sup>15</sup> offre un bon exemple de graphe de connaissance ouvert grâce auquel une information lisible aussi bien par les machines que par les humains s'accumule progressivement. Néanmoins, chacun de ces graphes de connaissance est organisé selon les finalités – toujours particulières – de ses auteurs, et ne peut être réutilisable facilement pour d'autres fins. Ni l'IA statistique, ni l'IA symbolique ne possèdent les propriétés de *recombinaison fluide* que l'on est en droit d'attendre des modules d'une intelligence artificielle au service de l'intelligence collective.

### **L'IA symbolique est gourmande en travail intellectuel humain**

On a bien tenté d'enfermer toute la connaissance humaine dans une seule ontologie afin de permettre une meilleure interopérabilité, mais alors la richesse, la complexité, l'évolution et les multiples perspectives du savoir humain sont effacées. Sur un plan pratique, les ontologies universelles – voire celles qui prétendent formaliser l'ensemble des catégories, relations et règles logiques d'un vaste domaine – deviennent vite énormes, touffues, difficiles à comprendre et à maintenir pour l'humain qui est amené à s'en occuper. Un des principaux goulets d'étranglement de l'IA symbolique est d'ailleurs la quantité et la haute qualité du *travail humain* nécessaire à modéliser un domaine de connaissance, aussi étroitement circonscrit soit-il. En effet, il est non seulement nécessaire de lire la documentation, mais il faut encore interroger et écouter longuement plusieurs experts du domaine à modéliser. Acquis par l'expérience, les savoirs de ces experts s'expriment le plus souvent par des récits, des exemples et par la description de situations-types. Il faut alors transformer une connaissance empirique de style oral en un modèle logique cohérent dont les règles doivent être exécutables par un ordinateur. En fin de compte, le raisonnement des experts sera bien automatisé, mais le travail "d'ingénierie de la connaissance" d'où procède la modélisation ne peut pas l'être.

<sup>15</sup> [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page).

## Position du problème: quel est le principal obstacle au développement de l'IA?

### Vers une intelligence artificielle neuro-symbolique

Il est maintenant temps de prendre un peu de recul. Les deux branches de l'IA – neuronale et symbolique – existent depuis le milieu du XXe siècle et elles correspondent à deux styles cognitifs également présents chez l'humain. D'une part, nous avons la reconnaissance de formes (*pattern recognition*) qui correspond à des modules sensorimoteurs réflexes, que ces derniers soient appris ou d'origine génétique. D'autre part, nous avons une connaissance conceptuelle explicite et réfléchie, souvent organisée en modèles causaux et qui peut faire l'objet de raisonnements. Comme ces deux styles cognitifs fonctionnent ensemble dans la cognition humaine, il n'existe aucune raison théorique pour ne pas tenter de les faire coopérer dans des systèmes d'intelligence artificielle. Les bénéfices sont évidents et, en particulier, chacun des deux sous-systèmes peut remédier aux problèmes rencontrés par l'autre. Dans une IA mixte, la partie symbolique surmonte les difficultés de conceptualisation, de généralisation, de modélisation causale et de transparence de la partie neuronale. Symétriquement, la partie neuronale amène les capacités de reconnaissance de forme et d'apprentissage à partir d'exemples qui font défaut à l'IA symbolique.

Aussi bien d'importants chercheurs en intelligence artificielle que de nombreux observateurs avertis de la discipline poussent dans cette direction d'une IA hybride. Par exemple, Dieter Ernst a récemment défendu une "*intégration* entre les réseaux neuronaux, qui excellent dans la classification des perceptions et les systèmes symboliques, qui excellent dans l'abstraction et l'inférence"<sup>16</sup>. Emboîtant le pas à Gary Marcus, les chercheurs en IA Luis Lamb et Arthur D'avila Garcez ont récemment publié un article<sup>17</sup> en faveur d'une IA neuro-symbolique dans laquelle des représentations acquises par des moyens neuronaux seraient interprétées et traitées par des moyens symboliques. Il semble donc que l'on ait trouvé une solution au problème du blocage de l'IA: il suffirait d'accoupler intelligemment les branches symbolique et statistique plutôt que de les maintenir séparées comme deux programmes de recherche en concurrence<sup>18</sup>.

<sup>16</sup> D. Ernst, *AI Research and Governance Are at a Crossroads*: <https://www.cigionline.org/articles/ai-research-and-governance-are-crossroads/>.

<sup>17</sup> A. D'Avila Garcez, L. Lamb, *Neurosymbolic AI: The 3rd Wave*, in "ArXiv", Décembre, 2020.

<sup>18</sup> Voir le récent rapport de l'Université de Stanford "100 Year Study on AI" qui identifie le courant neuro-symbolique comme une des clés de l'avancement de la discipline: <https://ai100.stanford.edu/>.

D'ailleurs, ne voit-on pas les grandes compagnies du Web, qui mettent en avant l'apprentissage automatique et l'IA neuronale dans leurs efforts de relations publiques, développer plus discrètement en interne des *graphes de connaissance* pour organiser leur mémoire numérique et donner sens aux résultats des réseaux neuronaux? Mais avant de déclarer la question réglée, réfléchissons encore un peu aux données du problème.

### **Cognition animale et cognition humaine**

Pour chacune des deux branches de l'IA, nous avons dressé une liste des obstacles qui se dressent sur le chemin menant vers une intelligence artificielle moins fragmentée, plus utile et plus transparente. Or nous avons trouvé un même inconvénient des deux côtés: le cloisonnement logique, les difficultés d'accumulation et d'intégration. Réunir le neuronal au symbolique ne nous aidera pas à surmonter cet obstacle puisque ni l'un ni l'autre n'en sont capables. Pourtant, les sociétés humaines réelles peuvent transformer des perceptions muettes et des savoir-faire issus de l'expérience en connaissances partageables. À force de dialogue, un spécialiste d'un domaine finit par se faire comprendre d'un spécialiste d'un autre domaine et va peut-être même lui enseigner quelque chose. Comment reproduire ce type de performances cognitives dans des sociétés de machines? Qu'est-ce qui joue le rôle intégrateur du *langage naturel* dans les systèmes d'intelligence artificielle?

Bien des gens pensent que, le cerveau étant le support organique de l'intelligence, les modèles neuronaux sont la clé de sa simulation. Mais de quelle intelligence parle-t-on? N'oublions pas que *tous les animaux* ont un cerveau, or ce n'est pas l'intelligence du moucheron ou de la baleine que l'IA veut simuler, mais celle de l'humain. Et si nous sommes "plus intelligents" que les autres animaux (au moins de notre point de vue) ce n'est pas à cause de la taille de notre cerveau. L'éléphant possède un plus gros cerveau que l'Homme en termes absolus et le rapport entre la taille du cerveau et celle du corps est plus grand chez la souris que chez l'humain. C'est principalement notre *capacité linguistique*, notamment supportée par les aires de Broca, Wernicke et quelques autres (uniques à l'espèce humaine), qui distingue notre intelligence de celle des autres vertébrés supérieurs. Or ces modules de traitement du langage ne sont pas fonctionnellement séparés du reste du cerveau, ils informent au contraire l'ensemble de nos processus cognitifs, y compris nos compétences techniques et sociales. Nos perceptions, nos actions, nos émotions et nos communications sont codées linguistiquement et notre mémoire est largement organisée par un système de coordonnées sémantiques fourni par le langage.

Fort bien, dira-t-on. Simuler les capacités humaines de traitement symbolique, y compris la faculté linguistique, n'est-ce pas précisément ce que

l'IA symbolique est censée faire? Mais alors comment se fait-il qu'elle soit cloisonnée en ontologies distinctes, qu'elle peine à assurer l'interopérabilité sémantique de ses systèmes et qu'elle ne parvienne si difficilement à accumuler et à échanger les connaissances? Tout simplement parce que, malgré son nom de "symbolique", l'IA ne dispose toujours pas d'un modèle calculable du langage. Depuis les travaux de Chomsky<sup>19</sup>, nous savons calculer la dimension syntaxique des langues, mais leur dimension sémantique reste hors de portée de l'informatique. Afin de comprendre cette situation, il est nécessaire de rappeler quelques éléments de sémantique.

### La sémantique en linguistique

Du point de vue de l'étude scientifique du langage, la sémantique d'un mot ou d'une phrase se décompose en deux parties, mélangées dans la pratique, mais conceptuellement distinctes: la sémantique linguistique et la sémantique référentielle. En gros, la sémantique linguistique s'occupe des relations entre les mots alors que la sémantique référentielle traite de la relation entre les mots et les choses<sup>20</sup>.

La sémantique linguistique ou sémantique mot-mot. Un *symbole linguistique* (mot ou phrase) possède généralement deux faces: le *signifiant*, qui est une image visuelle ou acoustique et le *signifié* qui est un concept ou une catégorie générale. Par exemple, le signifiant "arbre", a pour signifié: "végétal ligneux, de taille variable, dont le tronc se garnit de branches à partir d'une certaine hauteur". La relation entre signifiant et signifié étant fixée par la langue, le signifié d'un mot ou d'une phrase se définit comme *un nœud de relations avec d'autres signifiés*. Dans un dictionnaire classique, chaque mot est situé par rapport à d'autres mots proches (le thésaurus) et il est expliqué par des phrases (la définition) utilisant des mots eux-mêmes expliqués par d'autres phrases, et ainsi de suite de manière circulaire. Un dictionnaire classique relève principalement de la *sémantique linguistique*. Les verbes et les noms communs (par exemple: arbre, animal, organe, manger) représentent des catégories qui sont elles-mêmes connectées par un dense réseau de *relations sémantiques* telles que: "est une partie de", "est un genre de", "appartient au même contexte que", "est la cause de", "est antérieur à", etc. Nous ne pouvons penser et communiquer à la manière humaine que parce que

<sup>19</sup> N. Chomsky, *Syntactic Structures*, The Hague and Paris, Mouton 1957; N. Chomsky, M. Schützenberger, *The Algebraic Theory of Context-Free Languages*, in *Computer Programming and Formal Languages*, P. Braffort, D. Hirschberg (a cura di), North Holland, Amsterdam 1963, pp. 118-161.

<sup>20</sup> P. Lévy, *Les fondements linguistiques d'IEML*, <https://pierrelevyblog.com/2021/01/21/les-fondements-linguistiques-dieml/>.

nos mémoires collectives et personnelles sont organisées par des catégories générales connectées par des relations sémantiques<sup>21</sup>.

La sémantique référentielle ou sémantique mot-chose. Par contraste avec la *sémantique linguistique*, la *sémantique référentielle* fait le pont entre un *symbole linguistique* (signifiant et signifié) et un *référent* (un individu réel). Lorsque je dis que “les platanes sont des arbres”, je précise le sens conventionnel du mot “platane” en le mettant en relation d’espèce à genre avec le mot “arbre” et je ne mets donc en jeu que la sémantique linguistique. Mais si je dis que “Cet arbre-là, dans la cour, est un platane”, alors je pointe vers un état de chose réel, et ma proposition est vraie ou fausse. Ce second énoncé met évidemment en jeu la sémantique linguistique puisque je dois d’abord connaître le sens des mots et la grammaire du français pour la comprendre. Mais s’ajoute à la dimension linguistique une sémantique référentielle puisque l’énoncé se rapporte à un objet particulier dans une situation concrète. Certains mots, comme les noms propres, n’ont pas de signifiés. Leur signifiant renvoie directement à un référent. Par exemple, le signifiant “Alexandre le Grand” désigne un personnage historique et le signifiant “Tokyo” désigne une ville. Par contraste avec un dictionnaire ordinaire, qui définit des concepts ou des catégories, un dictionnaire encyclopédique contient des descriptions *d’individus* réels ou fictifs pourvus de noms propres tels que divinités, héros de roman, personnages et événements historiques, objets géographiques, monuments, œuvres de l’esprit, etc. Sa principale fonction est de répertorier et de décrire des objets externes au système d’une langue. Il enregistre donc une sémantique référentielle<sup>22</sup>.

## La sémantique en IA

En informatique, les références ou individus réels (les réalités dont on parle) deviennent les *données* alors que les catégories générales deviennent les rubriques, champs ou *métadonnées* qui servent à classer et retrouver les données. Par exemple, dans la base de données d’une entreprise, “nom de l’employé”, “adresse” et “salaire” sont des catégories ou métadonnées tandis que “Tremblay”, “33 Boulevard René Lévesques” et “65 K\$ / an” sont des données. Dans ce domaine technique, la sémantique référentielle correspond au rapport entre données et métadonnées

<sup>21</sup> J’adopte ici en gros la position de Saul Kripke, suivie par la majorité des philosophes et grammairiens. Voir: S. Kripke, *Naming and Necessity*, Blackwell, Oxford 1980; tr. fr. *La logique des noms propres*, Minuit, Paris 1982.

<sup>22</sup> N.B.: Une *catégorie* est une classe d’individus, une abstraction. Il peut y avoir des catégories d’entités, de process, de qualités, de quantités, de relations, etc. Les mots “catégorie” et “concept” sont ici traités comme des synonymes.

et la sémantique linguistique au rapport entre les métadonnées ou catégories organisatrices, qui sont généralement représentées par des mots ou de courtes expressions linguistiques.

Dans la mesure ou la finalité de l'informatique est d'augmenter l'intelligence humaine, elle doit notamment nous aider à donner sens aux flots de données numériques et à en tirer le maximum de connaissances utiles pour l'action. À cet effet, nous devons catégoriser correctement les données – c'est-à-dire mettre en œuvre une sémantique mot-chose – et organiser les catégories selon des relations pertinentes, qui nous permettent d'extraire des données toutes les connaissances utiles pour l'action – ce qui correspond à la sémantique mot-mot.

En discutant le sujet de la sémantique en informatique, nous devons nous souvenir que les ordinateurs ne voient pas spontanément un mot ou une phrase comme un concept en relation déterminée avec d'autres concepts dans le cadre d'une langue, mais seulement comme des suites de lettres, des "chaînes de caractères". C'est pourquoi les relations entre les catégories qui semblent évidentes aux humains et qui relèvent de la sémantique linguistique, doivent être ajoutées – le plus souvent à la main – dans une base de données si l'on veut qu'un programme en tienne compte.

Examinons maintenant dans quelle mesure l'IA symbolique modélise la sémantique. Si l'on considère les ontologies du "Web Sémantique" (le standard en IA symbolique), on découvre que les sens des mots et des phrases n'y dépendent pas de la circularité auto-explicative de la langue (comme dans un dictionnaire classique), mais d'un renvoi à des URI (Uniform Resource Identifiers) *qui fonctionne sur le mode de la sémantique référentielle* (comme un dictionnaire encyclopédique). Au lieu de reposer sur des concepts (ou catégories) déjà donnés dans une langue et qui se présentent dès l'origine comme des nœuds de relations avec d'autres concepts, les échafaudages du Web sémantique s'appuient sur des concepts définis séparément les uns des autres au moyen d'identifiants uniques. La circulation du sens dans un réseau de signifiés est escamotée au profit d'une relation directe entre signifiant et référent, comme si tous les mots étaient des noms propres. En l'absence d'une sémantique linguistique fondée sur une grammaire et un dictionnaire commun, les ontologies restent donc cloisonnées. En somme, l'IA symbolique contemporaine n'a pas accès à la pleine puissance cognitive et communicative du langage parce qu'elle ne dispose pas d'une langue, mais seulement d'une sémantique référentielle rigide.

Pourquoi l'IA n'utilise-t-elle pas les langues naturelles – avec leur sémantique linguistique inhérente – pour représenter les connaissances? La réponse est bien connue: parce que les langues naturelles sont ambiguës. Un mot peut avoir plusieurs sens, un sens peut s'exprimer par plusieurs mots, les phrases ont plusieurs interprétations possibles, la grammaire

est élastique, etc. Comme les ordinateurs ne sont pas des êtres incarnés et pleins de bon sens, comme nous, ils ne sont pas capables de désambiguïser correctement les énoncés en langue naturelle. Pour ses locuteurs humains, une langue naturelle étend un filet de catégories générales prédéfinies qui s'expliquent mutuellement. Ce réseau sémantique commun permet de décrire et de faire communiquer aussi bien les multiples situations concrètes que les différents domaines de connaissance. Mais, du fait des limitations des machines, l'IA ne peut pas faire jouer ce rôle à une langue naturelle. C'est pourquoi elle reste aujourd'hui fragmentée en micro-domaines de pratiques et de connaissances, chacun d'eux avec sa sémantique particulière.

L'automatisation de la sémantique linguistique pourrait ouvrir de nouveaux horizons de communication et de raisonnement à l'intelligence artificielle. Pour traiter la sémantique linguistique, l'IA aurait besoin d'une langue standardisée et univoque, d'un code spécialement conçu à l'usage des machines, mais que les humains pourraient aisément comprendre et manipuler. Cette langue permettrait enfin aux modèles de se connecter et aux connaissances de s'accumuler. En somme, *le principal obstacle au développement de l'IA est l'absence d'un langage commun calculable*. C'est précisément le problème résolu par IEML<sup>23</sup>, qui possède la capacité d'exprimer le sens, comme les langues naturelles, mais dont la sémantique est non ambiguë et calculable, comme un langage mathématique. L'utilisation d'IEML rendra l'IA moins coûteuse en efforts humains, plus apte à traiter le sens et la causalité, et surtout, capable d'accumuler et d'échanger des connaissances.

## Une solution basée sur un codage de la sémantique

### Le métalangage de l'économie de l'information

Beaucoup de progrès en informatique viennent de l'invention d'un système de codage pertinent rendant l'objet codé (nombre, image, son, etc.) facilement calculable par une machine. Par exemple, le codage binaire pour les nombres et le codage en pixels ou en vecteurs pour les images.

<sup>23</sup> P. Lévy, *Toward a Self-referential Collective Intelligence: Some Philosophical Background of the IEML Research Program*, in N. Nguyen, R. Kowalczyk, C. Shyi-Ming (a cura di), *Computational Collective Intelligence, Semantic Web, Social Networks and Multiagent Systems*, First International Conference, ICCCI, Wroclaw (Poland), Oct. 2009, Springer, Berlin-Heidelberg-New York 2009, pp. 22-35; Id., *The IEML Research Program: From Social Computing to Reflexive Collective Intelligence*, in "Information Sciences, Special issue on Collective Intelligence", E. Kapetanios, G. Kourtrika (a cura di), 80, 1, Elsevier, Amsterdam 2010, pp. 71-94.

C'est pourquoi je me suis attaché à la conception d'un code qui rende la sémantique linguistique calculable. Cette langue artificielle, IEML (Information Economy MetaLanguage)<sup>24</sup> possède une grammaire régulière et un dictionnaire compact de trois mille mots. Des catégories complexes peuvent être construites en combinant les mots selon les règles de la grammaire. Les catégories complexes peuvent à leur tour être utilisées pour en définir d'autres, et ainsi de suite récursivement. Une des parties les plus difficiles de la conception d'IEML a été de trouver le plus petit ensemble de mots à partir duquel n'importe quelle catégorie pourrait être construite.

Sur un plan linguistique, IEML possède la même capacité expressive qu'une langue naturelle. Elle peut donc traduire n'importe quelle autre langue. C'est d'autre part une langue univoque: ses expressions n'ont qu'un seul sens. Enfin, sa sémantique est calculable. Cela signifie que son dictionnaire et ses règles de grammaire suffisent à déterminer automatiquement le sens de ses expressions (ce qui n'est pas le cas des langues naturelles). Soulignons qu'IEML n'est pas une ontologie universelle, mais bel et bien une *langue* qui permet d'exprimer n'importe quelle ontologie ou classification particulière.

Sur un plan mathématique, IEML est un langage régulier au sens de Chomsky: c'est une algèbre. Elle est donc susceptible de toutes sortes de traitements et de transformations automatiques.

Sur un plan informatique, comme nous le verrons plus en détail ci-dessous, le métalangage donne prise à un langage de programmation de réseaux sémantiques et supporte le système d'indexation d'une base de connaissances<sup>25</sup>.

## L'éditeur IEML

Le métalangage de l'économie de l'information est défini par sa grammaire et son dictionnaire, que l'on trouvera en consultant le site [intlekt.io](https://intlekt.io). Mais la *langue* ne suffit pas. Nous avons besoin d'un *outil numérique* facilitant son écriture, sa lecture et son utilisation: l'éditeur IEML.

L'éditeur IEML sert à produire et à explorer des *modèles de données*. Cette notion de "modèle" englobe les réseaux sémantiques, les systèmes

<sup>24</sup> Sur la grammaire d'IEML, voir: <https://intlekt.io/ieml-grammar/>. Sur le dictionnaire d'IEML, voir: <https://intlekt.io/ieml-dictionary/>.

<sup>25</sup> Autre références pertinentes: P. Lévy, *L'intelligence collective, pour une anthropologie du cyberspace*, La Découverte, Paris 1994; Id., *Les systèmes à base de connaissance comme médias de transmission de l'expertise (knowledge based systems as media for transmission of expertise)*, in "Intellectica special issue on Expertise and cognitive sciences", Violaine Prince (a cura di), 12, 1991. pp. 187-219; Id., *De la programmation considérée comme un des beaux-arts*, La Découverte, Paris 1992.

de métadonnées sémantiques, les ontologies, les graphes de connaissances et les systèmes d'étiquettes pour catégoriser des données d'entraînement. L'éditeur contient un langage de programmation permettant d'automatiser la création de nœuds (les catégories) et de liens (les relations sémantiques entre catégories). Ce langage de programmation est de type *déclaratif*, c'est-à-dire qu'il ne demande pas à son utilisateur d'organiser des flots d'instructions conditionnelles, mais seulement de décrire les résultats à obtenir.

### Mode d'utilisation de l'éditeur

Comment se sert-on de l'éditeur?

1. Le modélisateur répertorie les *catégories* qui vont servir de conteneurs (ou de cases-mémoire) aux différents types de données. S'il a besoin de catégories qui ne correspondent à aucun des 3000 *mots* du dictionnaire IEML il les crée au moyen de *phrases*.

2. Il programme ensuite les *relations sémantiques* qui vont connecter les données catégorisées. Les relations, définies par des phrases, ont un contenu sémantique aussi varié que nécessaire. Leurs propriétés mathématiques (réflexivité, symétrie, transitivité) sont spécifiées. Des *instructions* conditionnent l'établissement des relations à la présence de signifiants ou de valeurs de données déterminées à certaines adresses syntaxiques des catégories.

3. Une fois les données catégorisées, le programme tisse automatiquement le réseau de relations sémantiques qui va leur donner sens. Des fonctions de fouille de données, d'exploration hypertextuelle et de visualisation des relations par tables et par graphes permettent aux utilisateurs finaux d'explorer le contenu modélisé.

### Avantages

Plusieurs traits fondamentaux distinguent l'éditeur IEML des outils contemporains qu'on utilise pour modéliser les données: les catégories et relations sont programmables, les modèles obtenus sont interopérables et transparents.

Catégories et relations sont programmables. La structure régulière d'IEML permet de générer les catégories et de tisser les relations de manière *fonctionnelle ou automatique* au lieu de les créer une par une. Cette propriété fait gagner au modélisateur un temps considérable. Le temps gagné par l'automatisation de la création des catégories et des relations compense largement le temps passé à coder les catégories en IEML, et cela d'autant plus qu'une fois créées, les nouvelles catégories et relations peuvent être échangées entre les utilisateurs.

Les modèles sont interopérables. *Tous les modèles se ramènent au même dictionnaire de trois mille mots et à la même grammaire.* Les modèles sont donc *interopérables*, c'est-à-dire qu'ils peuvent facilement fusionner ou échanger des catégories et des sous-modèles. Chaque modèle reste adapté à un contexte particulier, mais les modèles peuvent désormais se comparer, s'interconnecter et s'intégrer.

Les modèles sont transparents. *Bien qu'ils soient codés en IEML, les modèles rédigés au moyen de l'éditeur IEML sont lisibles en langue naturelle.* De plus, les catégories et relations se présentent comme des mots ou des phrases. Or les mots sont expliqués par leurs relations avec les autres mots du dictionnaire et le sens des phrases est défini par les mots qui les composent selon une grammaire régulière. *Toutes les catégories et toutes les relations sont donc explicitement définies*, ce qui rend les modèles *clairs* aussi bien pour les modélisateurs que pour les utilisateurs finaux et adéquats aux *principes d'éthique et de transparence* contemporains.

Au prix d'un bref apprentissage, l'éditeur peut être utilisé par des non-informaticiens qui ne connaissent pas la langue IEML. Seule la grammaire (simple et régulière) doit être maîtrisée, les mots IEML étant représentés en langues naturelles. L'éditeur IEML pourrait être utilisé dans les écoles et ouvrir la voie à une démocratisation de la maîtrise des données.

### L'architecture neuro-sémantique

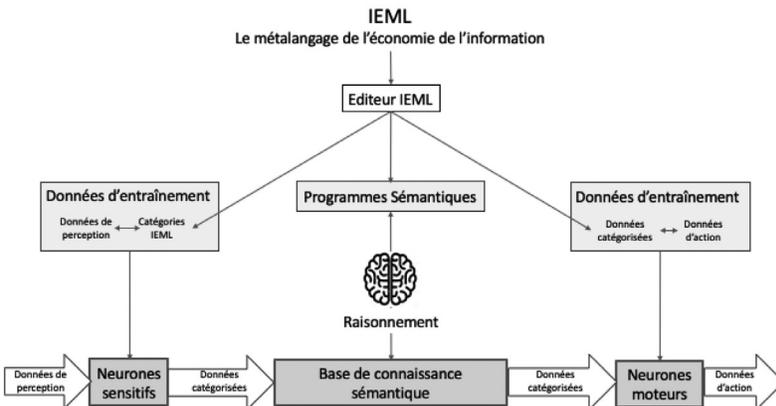


Figure 1: Une architecture Neuro-sémantique pour l'IA

Je vais maintenant proposer une architecture de système d'IA basée sur IEML. Cette architecture (schématisée dans la figure 1) est évidemment un cas particulier d'architecture neuro-symbolique, mais je la nomme *neuro-sémantique* afin de souligner qu'elle résout le problème du *calcul de la sémantique* et de *l'interopérabilité sémantique* entre systèmes.

### Les neurones sensorimoteurs

Le module d'entrée est occupé par des réseaux de neurones sensoriels, qui ont été entraînés par des exemples de données catégorisées en IEML. On doit distinguer plusieurs types de données d'entraînement (texte, image, sons, etc.) d'où résultent plusieurs types de réseaux de neurones. Les données catégorisées par les neurones sensoriels sont transmises à la base de connaissance sémantique. Si l'on détecte des incohérences, des erreurs ou des biais, il faut évidemment revoir les données d'entraînement ou réviser leur conceptualisation. Le système doit donc comprendre une boucle de dialogue entre les annotateurs de données qui entraînent les réseaux de neurones et les ingénieurs qui gèrent la base de connaissance.

En sortie, des réseaux de neurones moteurs transforment des données catégorisées en données qui commandent des actions, telles que rédaction de texte, synthèse d'image, émission vocale, instructions envoyées à des effecteurs (robots), etc. Ces neurones moteurs sont entraînés sur des exemples qui appartiennent à des données catégorisées en IEML et des données motrices. Là encore, les données d'entraînement et les réseaux de neurones doivent être distingués selon leurs types.

### La mémoire et le traitement sémantique

La base de connaissance est organisée par un réseau sémantique. Elle est donc de préférence supportée par une base de données de graphes (*graph database*). Sur le plan de l'interface, cette base de connaissance se présente comme une encyclopédie hypertextuelle. Elle autorise aussi la programmation de simulations et de divers tableaux de bord pour la veille et le renseignement.

L'éditeur IEML évoqué à la section précédente peut servir à d'autres tâches qu'à la modélisation. Il permet en effet de conditionner les opérations d'écriture-lecture les plus variées à la présence de contenus sémantiques situés à certaines adresses syntaxiques. Lorsqu'ils sont codés en IEML les concepts deviennent les variables d'une algèbre, ce qui n'est évidemment pas le cas lorsqu'elles sont exprimées en langue naturelle. C'est pourquoi des *transformations sémantiques* peuvent être programmées et calculées. Cette programmation sémantique ouvre la voie non seulement aux *raisonnements logiques* classiques auxquels les moteurs

d'inférence de l'IA symbolique nous ont habitués depuis des décennies, mais aussi à d'autres formes de raisonnement automatique. Puisqu'en IEML la sémantique est une image fonctionnelle de la syntaxe, il devient possible d'automatiser le *raisonnement analogique* de type "A est à B ce que C est à D". D'autres d'opérations sémantiques peuvent également être programmées, telles que sélection et fouille; substitution, insertion ou effacement; extraction de sous-réseaux sémantiques pertinents; résumé ou développement; inversion, allusion, atténuation ou amplification; extraction ou projection de structures narratives, et ainsi de suite.

### Quelques applications

Quelques applications évidentes de notre architecture d'IA neuro-sémantique sont: l'intégration de données, l'aide à la décision à partir de modèles causaux, la gestion des connaissances, la compréhension et le résumé de texte, la génération de texte *contrôlée* (contrairement aux systèmes de type GPT-3 dont le texte n'est pas contrôlé), les chatbots et la robotique. Je vais maintenant brièvement commenter deux exemples d'usage: la compréhension de texte et la génération de texte contrôlée.

Concernant la *génération de texte contrôlée*, imaginons en entrée des données de télémétrie, des informations comptables, des examens médicaux, des résultats de tests de connaissance, etc. On peut alors concevoir en sortie des textes narratifs en langue naturelle synthétisant le contenu des flux de données d'entrée: diagnostics médicaux, bulletins scolaires, rapports, conseils, etc. Quant à la *compréhension de texte*, elle suppose d'abord la catégorisation automatique du contenu du document présenté en entrée du système. Dans un deuxième temps, le modèle sémantique extrait du texte est inscrit dans la mémoire du système de manière à s'intégrer aux connaissances déjà acquises. En somme, des systèmes d'intelligence artificielle pourraient accumuler des connaissances à partir de la lecture automatique de documents. À supposer qu'IEML soit adopté, les systèmes d'intelligence artificielle deviendraient non seulement capables d'accumuler des connaissances, mais de les intégrer en modèles cohérents et de les échanger. Il s'agit évidemment là d'une perspective à long terme qui exigera des efforts coordonnés.

### Conclusion: vers un tournant humaniste en IA

Sans langage, nous n'aurions accès ni au questionnement, ni au dialogue, ni au récit. La langue est simultanément un médium de l'intelligence personnelle – il est difficile de penser sans dialogue intérieur – et de l'intelligence collective. La plupart de nos connaissances ont été accumulées

et transmises par la société sous forme linguistique. Vu le rôle de la parole dans l'intelligence humaine, Il est surprenant qu'on ait espéré atteindre une intelligence artificielle générale sans disposer d'un modèle calculable du langage et de sa sémantique. La bonne nouvelle est que nous en avons finalement un. Même si l'architecture neuro-sémantique ici proposée ne débouche pas directement sur une intelligence artificielle générale, elle autorise au moins la construction d'applications capables de traiter le *sens* des textes ou des situations. Elle permet aussi d'envisager un marché des données privées étiquetées en IEMML qui stimulerait, s'il en était besoin, le développement de l'apprentissage statistique. Elle devrait aussi supporter une mémoire publique collaborative qui serait particulièrement utile dans les domaines de la recherche scientifique, de l'éducation et de la santé.

La multiplicité des langues, des systèmes de classification, des points de vue disciplinaires et des contextes pratiques cloisonnent aujourd'hui la mémoire numérique. Or la communication des modèles, la comparaison critique des points de vue et l'accumulation des connaissances sont essentielles à la cognition symbolique humaine, une cognition indissolublement personnelle et collective. L'intelligence artificielle ne pourra durablement augmenter la cognition humaine qu'à la condition d'être interopérable, cumulable, intégrable, échangeable et distribuée. C'est dire qu'on ne fera pas de progrès notable en intelligence artificielle sans viser en même temps une intelligence collective capable de se réfléchir et de se coordonner dans la mémoire mondiale. L'adoption d'une langue calculable fonctionnant comme système universel de coordonnées sémantiques – une langue facile à lire et à écrire permettant de tout dire comme de distinguer les nuances – ouvrirait de nouvelles voies à l'intelligence collective humaine, y compris sous la forme d'une interaction immersive multimédia dans le monde des idées. En ce sens, la communauté des utilisateurs d'IEMML pourrait inaugurer une nouvelle époque de l'intelligence collective.

L'IA contemporaine, majoritairement statistique, a tendance à créer des situations où les données pensent à notre place et à notre insu. Par contraste, je propose de développer une IA qui aide les humains à prendre le contrôle intellectuel des données pour en extraire un sens partageable de manière durable. IEMML nous permet de repenser la finalité et le mode d'action de l'IA d'un point de vue humaniste, point de vue pour qui le sens, la mémoire et la conscience personnelle doivent être traités avec le plus grand sérieux.

## Bibliographie

Berners-Lee T., *Weaving the Web. The original design and ultimate destiny of the world wide web by its inventor*, Harper, San Francisco 1999.

- Bush V., *As We May Think*, Atlantic Monthly, Boston 1945.
- Chomsky N., *Syntactic Structures*, The Hague and Paris, Mouton 1957.
- Id., *New Horizons in the Study of Language and Mind*, Cambridge UP 2000.
- Chomsky N., Schützenberger M., *The Algebraic Theory of Context-Free Languages*, in *Computer Programming and Formal Languages*, P. Braffort, D. Hirschberg (a cura di), North Holland, Amsterdam 1963, pp. 118-161.
- D'Avila Garcez A., Lamb L., *Neurosymbolic AI: The 3rd Wave*, in "arXiv", Cornell University 2020.
- Engelbart D., *Augmenting Human Intellect. Technical Report*, Stanford Research Institute, Stanford 1962.
- Ernst D., *AI Research and Governance Are at a Crossroads*, in <https://www.cigionline.org/articles/ai-research-and-governance-are-crossroads/>
- Kripke S., *Naming and Necessity*, Blackwell, Oxford 1980; tr. fr. *La logique des noms propres*, Minuit, Paris 1982.
- Lévy P., *L'Oeuvre de Warren McCulloch*, in "Cahiers du CREA", 7, Paris 1986, pp. 211-255.
- Id., *Analyse de contenu des travaux du Biological Computer Laboratory (BCL)*, in "Cahiers du CREA", 8, Paris, 1986, pp. 155-191.
- Id., *Les systèmes à base de connaissance comme médias de transmission de l'expertise* (knowledge based systems as media for transmission of expertise), in "Intellectica special issue on "Expertise and cognitive sciences", Violaine Prince (a cura di), 12, 1991. pp. 187 – 219.
- Id., *De la programmation considérée comme un des beaux-arts*, La Découverte, Paris, 1992.
- Id., *L'intelligence collective, pour une anthropologie du cyberspace*, La Découverte, Paris 1994; tr. ingl. di R. Bonono, *Collective Intelligence*, Perseus Books, Cambridge MA 1997.
- Id., *Toward a Self-referential Collective Intelligence: Some Philosophical Background of the IEML Research Program*, in N. Nguyen, R. Kowalczyk, C. Shyi-Ming (a cura di), *Computational Collective Intelligence, Semantic Web, Social Networks and Multiagent Systems, First International Conference, ICC-CI, Wroclaw (Poland), Oct. 2009*, Springer, Berlin-Heidelberg-New York 2009, pp. 22-35.
- Id., *The IEML Research Program: From Social Computing to Reflexive Collective Intelligence*, in "Information Sciences, Special issue on Collective Intelligence", E. Kapetanios, G. Koutrika (a cura di), 80, 1, Elsevier, Amsterdam 2010, pp. 71-94.
- Id., *La Sphère sémantique. Computation, cognition, économie de l'information*, Hermes-Lavoisier, Paris / Londres 2011; tr. ingl. *The Semantic Sphere. Computation, cognition and information economy*, Wiley, London e NY 2011.
- Id., *Les fondements linguistiques d'IEML*, in <https://pierrelevyblog.com/2021/01/21/les-fondements-linguistiques-dieml/>.
- Id., *Esquisse d'un modèle d'affaire pour un changement de civilisation*, in <https://pierrelevyblog.com/2021/04/03/esquisse-dun-modele-daffaire-pour-un-changement-de-civilisation/>.
- Id., *Les noms propres en IEML*, in <https://pierrelevyblog.com/2021/07/13/les-noms-propres-en-ieml/>.

- Licklider J., *Man-Computer Symbiosis*, in “IRE Transactions on Human Factors in Electronics”, 1, 1960, pp. 4-11.
- Marcus G., *Deep learning: a critical appraisal*, in “arXiv”, Cornell University 2018.
- Marcus G., Davis E., *Rebooting AI: Building Artificial Intelligence We Can Trust*, Vintage, New York 2019.
- McClelland J., Rumelhart D., PDP research group (a cura di), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 2 Voll., MIT Press, Cambridge MA 1986.
- McCulloch W., *Embodiments of Mind*, MIT Press, Cambridge MA 1965.
- McCulloch W., Pitts W., *A Logical Calculus of Ideas Immanent in Nervous Activity*, in “Bulletin of Mathematical Biophysics”, 5, 1943, pp. 115-133.
- Minsky M., Papert S., *Perceptron*, MIT Press, Cambridge MA 1969.
- Minsky M., *The Society of Mind*, Simon & Schuster, New York 1986.
- Pearl J., Mackenzie D., *The Book of Why. The new science of cause and effect*, Basic books, New York 2019.
- Popper K., *Objective Knowledge: An Evolutionary Approach*, Clarendon Press, Oxford 1972.
- Rosenblatt F., *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan Books, Washington 1962.
- Rumelhart D., Geoffrey E., Hinton G., Williams R., *Learning representations by back-propagating errors*, in “Nature”, 323, 1985, pp. 533-536.
- Von Foerster H., *Observing Systems: Selected Papers of Heinz von Foerster*, Intersystems Publications, Seaside CA 1981.