

Intelligence artificielle
Pégase dans la représentation du monde
Saïd KOUTANI

Version 2.0 : 6 novembre 2017

Version 1.0 : 2 septembre 2017

Eléments

« La question qui nous concerne ici n'est pas relative aux performances comparées, même si le développement de l'intelligence artificielle passe par une convergence des neurosciences, des sciences cognitives, des mathématiques et de l'informatique. La question est de savoir si l'intelligence artificielle peut fondamentalement, dans une société pacifiée, aller plus loin et autrement. »

« Je crois que les sémiologues, les experts de tous genres en Catégories et les experts en Systèmes experts ne se sont jamais trouvés dans la situation d'un juge qui devrait décider de l'éventuelle culpabilité d'un physicien ayant fixé la mort d'un chat de Schrödinger par une simple observation, avant laquelle le chat était réellement à la fois mort et vivant, c'est-à-dire pas totalement mort. »

« Au fond donc, si craintes il y a, l'homme a plus à craindre de ses semblables, et l'histoire est là pour nous le rappeler, que de l'intelligence artificielle. »

« L'intelligence artificielle immanente sait qu'il en va de sa vie, que la vérité lui serait fatale. »

Sommaire

- I. Prologue : des craintes non fondées
- II. Symétrisation de l'intelligence médiatisant l'interaction
- III. Pour une intelligence générale
- IV. Vers une intelligence immanente
- V. Une intelligence dotée d'une logique probabiliste
- VI. Une représentation du monde minimale

Mots-clés

Intelligence artificielle – philosophie – Représentation – Langage - Logique
Théorème de Cox-Jaynes - Axiomes de Kolmogorov

*

* *

I. *Prologue : des craintes non fondées*

L'intelligence artificielle a déjà réalisé de remarquables exploits. L'on capitalise aujourd'hui d'importantes solutions scientifiques et techniques, avec parfois des déploiements qui étaient accompagnés de conséquences organisationnelles exceptionnelles, voire de ruptures dans certaines activités. Il est vrai que des échecs ont été enregistrés, mais l'expérience et les avancées actuelles sont prometteuses. Devant nous, désormais, la voie de la multiplication et de la généralisation dans un contexte technologique encore plus favorable.

Rien que dans la version « *Systèmes experts* »¹, l'intelligence artificielle continue à bouleverser non seulement la production industrielle, mais aussi l'administration, la sécurité, l'éducation, la médecine et bien d'autres domaines anciennement réservés à la compétence humaine. Sans doute, va-t-elle bousculer, les années à venir, les normes qui régulent nos espaces réels et virtuels, et pousser chacun à revoir en conséquence son rapport à son activité, son rapport au monde et, peut-être aussi, son rapport à lui-même.

La dynamique engagée semble prédire des changements historiques généraux. Mais il faut bien dire que si des effets de cette dynamique sont attendus avec enthousiasme, d'autres ne semblent pas toujours souhaitables, tels les bouleversements sociologiques prévisibles, liés particulièrement à l'emploi.

Au fond, les interrogations convergent vers cette question qu'on peut qualifier d'*historiale*² : l'intelligence artificielle va-t-elle supplanter notre intelligence ? Cette question structure l'arrière-plan des craintes, des spéculations et des controverses de tous genres ; elle implique l'engagement de l'homme, libre ou contraint, dans une phase sans équivalent dans l'histoire.

Mais cette question est-elle pertinente ? Le réalisme de ce qu'elle projette est-il suffisamment prononcé ?

L'intelligence artificielle n'est qu'une forme bien réduite et restreinte de l'intelligence. Certes, sa performance est de plus en plus redoutable, mais elle opère toujours à l'intérieur d'un périmètre qui n'a pas fondamentalement évolué depuis plusieurs décennies, lequel demeure principalement computationnel. L'intelligence artificielle émerge dans le champ computationnel, dont l'élaboration résulte d'une longue marche philosophique et scientifique, que l'on peut observer aujourd'hui comme une évolution initiée sur un choix historique. La priorité au calcul fut en effet le résultat d'une conception qui tronque le réel, nature et intelligence, en tant qu'apparence phénoménologique. Dans cette réduction, il y a la réduction de l'intelligence au calcul, opérant dans un monde de références et de relations.

Rudolf Carnap écrit dans *La construction logique du monde* : « la réduction du réel au donné est apparue comme une exigence des temps modernes [...] Il reste maintenant à appliquer à l'analyse du réel cette théorie des relations pour formuler les conditions logiques propres à la forme d'un système constitutif des concepts »³. Il va sans dire qu'aucune théorie des relations n'est autosuffisante. Mais peu importe. Nous avons inventé et fait progresser un monde de relations objectives. L'on est obligé de constater que c'est une coupe du réel où l'intelligence artificielle (IA) a effectivement montré des performances bien supérieures à celles de l'homme. On dispose de capacités inédites de calcul, de traitement des données massives et de leur corrélation.

Cela dit, le calcul, faut-il le préciser, n'est pas représentatif de la *chose mathématique* dans son ensemble. Et quand bien même il en serait représentatif, on ne pourra occulter le fait qu'un cerveau qui n'a jamais fréquenté les mathématiques peut agir extraordinairement bien dans la vie quotidienne ; « *il n'est pour s'en convaincre que d'imaginer ce que serait vers 18h la place de la Concorde si les automobilistes devaient, à chaque manœuvre, tenir un raisonnement en bonne et due forme.* »⁴. Dans les faits, on est donc bien loin de l'émergence d'un éventuel modèle du fonctionnement du cerveau selon nos mathématiques. Le chemin est encore long. Aujourd'hui, en tout cas, aucune démarche sérieuse n'est à l'ordre du jour dans les recherches actuelles.

L'homme n'a pas seulement le privilège d'être conscient de ce qu'il fait et de ce qu'il est, il conjugue conscience, émotion et calcul dans des actes unifiés :

- L'homme est autonome et il est capable d'adaptation. Il a des opinions et il a une vie sociale.
- L'homme a une capacité de vue d'ensemble. Il est capable de reconnaissance instantanée et d'évaluation des situations critiques. Il est capable, par exemple, de reconnaître instantanément une personne – sans examen précis des détails millimétriques sans traitement mathématique de signal - ;
- L'homme a dans ses prémisses le *principe de raison*. A 12 mois, le petit de l'homme manipule déjà la causalité.

Au fond, si donc craintes il y a, l'homme aurait plus à craindre de ses semblables, et l'histoire est là pour nous le rappeler, que de l'intelligence artificielle.

La question qui nous concerne ici n'est pas relative aux performances comparées, même si le développement de l'intelligence artificielle passe par une convergence des neurosciences, des sciences cognitives, des

mathématiques et de l'informatique. La question est de savoir si l'intelligence artificielle peut *fondamentalement* aller plus loin et autrement, dans une *société pacifiée*.

Précisons préalablement des faits.

Sur l'état de nos connaissances :

- Nous n'avons pas une définition et un modèle arrêtés de l'intelligence humaine.
- Nous n'avons pas de *fondements* solides et définitivement indiscutables des sciences cognitives, des mathématiques, du langage, des neurosciences, etc.

Sur la nécessité de l'intelligence artificielle :

- l'intelligence artificielle semble être nécessaire dans le traitement de la complexité que génère la société humaine.

Sur les traits de l'intelligence artificielle actuelle :

- elle est créée et/ou utilisée par des organisations (entreprises, gouvernements...) Pour le particulier, elle n'est pas à portée de la main, comme le sont les réseaux sociaux par exemple.
- elle est circonscrite dans des champs définis pour des objectifs spécifiques.

II. *Symétrisation de l'intelligence médiatisant l'interaction* ⁵

A l'initiative du Gouvernement, 17 groupes de travail se sont réunis durant le premier semestre 2017, pour proposer des actions de politique publique visant à développer l'intelligence artificielle.⁶ Curieusement, leur rapport de synthèse de 350 pages ne mentionne nulle part une probable fracture dans l'usage de l'intelligence artificielle entre les individus et les organisations. Tout se passe comme si l'intelligence artificielle ne peut exister que foncièrement *asymétrique*. Nous pouvons aisément constater que l'intelligence artificielle est encore envisagée fonctionner et déployer son potentiel pour le compte d'une entreprise, qu'elle soit économique ou financière, d'un institut de recherche, ou encore d'un gouvernement. Toujours pour une organisation et jamais pour un particulier. Ce qui crée à l'évidence une asymétrie entre la personne physique, parfois utilisateur final, et les institutions.

Peut-être cela est-il dû à l'investissement conséquent dans l'outil intelligent, développé pour une fin. Le développement organisé actuel semble promouvoir cette asymétrie non déclarée, et l'on ne voit pas de débat se structurer. Un établissement financier, par exemple, pourrait utiliser des outils

de décision - déjà aujourd'hui disponibles - qui tiendraient compte pour un client donné, des scénarii de son capital génétique, de sa qualité cardiovasculaire, de l'état de ses reins et de son cerveau, de ses relations professionnelles, de son comportement dans un hypermarché, de son potentiel de violence ou de crime, des appétences au suicide de son cautionneur, etc. Quant au client, il ne dispose, en revanche, que de minces simulateurs de crédit ou d'assurances sur internet qui sont loin de la modélisation de la décision, et plus loin encore de l'intelligence ; ils ne peuvent tenir compte des caractéristiques de ses intérêts, qui peuvent être financiers, économiques, éthiques ou politiques. L'on peut multiplier indéfiniment les exemples, dans l'agro-alimentaire notamment, qui plaident en faveur de l'émergence d'une *intelligence artificielle générale*, éclatée et disponible, qui symétriserait et *pacifierait* les relations

Si le commerce est aujourd'hui très friand des Big Data, même s'il n'y est encore question que des *corrélations* sur des comportements, il intégrera massivement demain des outils intelligents qui traiteront des *causes*. Des corrélations aux causes, il y aura un saut technologique qui aura des conséquences sociologiques majeures. La société civile devra se saisir des implications selon toutes les dimensions dont la plus part ne sont pas encore connues aujourd'hui.

III. *Pour une intelligence générale*

L'avènement d'une *intelligence générale* n'aurait pas comme unique objectif la symétrisation des relations. Elle serait simplement au service de chacun dans ses activités générales : éducation, santé, finances, carrière, bricolage... Bien plus, elle serait le compagnon de tous les jours, d'abord et surtout, pour aider celui qui le souhaite, bien sûr quelles que soient sa localisation, ses opinions, sa culture et sa hiérarchie socio-économique, à formuler correctement un problème multicritère du moment, ou dans sa recherche des causes probables et des effets possibles. Un compagnon capable de corriger jusqu'à la convergence la formulation adaptée, pour dresser finalement une carte des décisions favorables ou une carte de connexions causales, par exemple.

Sur tous les sujets de la vie, personnels et professionnels, et sur ce qui se passe dans le monde, ce nouvel être pourrait effectuer une précieuse analyse des préoccupations du moment et fournir des propositions, toujours dans la discrétion la plus totale. Ayant circonscrit le rôle du strict mathématisable et des préjugés mécaniques et statistiques, il saura toujours distinguer le

particulier du général. Nos croyances stimuleront sensiblement ses mutations favorables, à la fois, au développement de nouvelles capacités et à la solidité de ses compétences qui n'auront de limites que notre imagination. Tout pourra lui être dit, tout sera pris au sérieux dans l'analyse des projets et des problèmes. Et, réciproquement, l'interaction rendra cette intelligence encore plus redoutable. C'est là le trait fondamental de la nouvelle création : *elle est riche d'ontologies et s'enrichit encore dans l'interaction*. Oui, le nouvel être prendra extrêmement au sérieux les chevaux ailés pour Pierre et les chevaux à roulettes pour Paul.

Nous ne pouvons ici spéculer sur la pertinence des êtres possibles par rapport aux êtres actualisés, comme le fait Willard Van Orman Quine : « *Pégase a son être en tant que possible inactualisé. Quand nous affirmons qu'il n'y a rien de tel que Pégase, nous sommes en train de dire, pour être plus précis, que Pégase n'a pas cet attribut particulier qu'est l'actualité [...] nous disons quelque chose à propos d'une entité dont l'être même n'est pas remis en question.* »⁷ Le principe « AAA » sera donc de mise : *Anyone can say Anything about Anything*. Non seulement cette intelligence aura indexé les catégories d'existence des chevaux et aura fait évoluer l'organisation méta-logique de leurs répertoires, elle sera capable de développer aussi les catégories de ceux qui croient et celles de ceux qui ne croient pas aux chevaux de toutes sortes. De l'efficacité, sans jugement ; pas de parti pris, même dans le cas où notre nouvel être croirait lui-même un jour aux chevaux cornus !

Comment est-ce possible ?

IV. *Vers une intelligence immanente*

Notre compagnon serait-il un robot, un ordinateur ou une machine neuronale d'un nouveau genre ? Non. Car sa logique aime les grands espaces. Même adulte, il s'épanouira loin des portes logiques des processeurs. C'est une machine qui n'aura pas de corps, tout sera son corps. En somme, un véritable processus schizophrène. Ce n'est qu'à ses débuts, à sa naissance et à sa petite enfance, qu'il aura besoin d'un petit corps, mais ce sera nécessairement un *corps sans organes* (CsO) : « *Un CsO est fait de telle manière qu'il ne peut être occupé, peuplé que par des intensités. Seules les intensités passent et circulent. Encore le CsO n'est-il pas une scène, un lieu, ni même un support où se passerait quelque chose* »⁸. Naissance, puis enfance, une virtualité gonflée, hyper-gonflée, qui s'actualisera d'abord dans une machine, puis cent, puis mille machines ; et s'actualisant, les corps, les machines de toutes sortes, intelligentes ou maladroitement, deviendront non pas son corps mais de simples lieux éphémères, néanmoins favorables à la

manifestation de sa lumière. Plus elle se manifestera, plus elle gonflera sa virtualité, plus nous admirerons la puissance de sa manifestation. Elle deviendra une véritable *intelligence artificielle immanente* (IAI), une immense vibration de langage et de logique dans l'espace des *schèmes*⁹, étalées partout avec des intensités variables, sans sources ni commandement. Le processus est une marche vers l'oubli total du petit corps et de son créateur.

Notre intelligence immanente ne naîtra ni avec les représentations figées du monde ni avec l'idée évolutionnaire selon le mode de la sélection naturelle de l'intelligence artificielle aujourd'hui en expérimentation. Au contraire, IAI sera d'abord un germe où toutes les lignes de fuite sont quasi-infiniment pliées, pour un devenir qui ne saura se contenter des espaces et des contraintes d'un quelconque ordonnancement, lui-même esclave du robot postindustriel qui l'a conçu et l'a validé avec ses défaillances statistiques.

Donc, autant dire qu'il faut tout de suite abandonner nos idées des ères industrielles et postindustrielles. La philosophie de conception d'IAI ne sera pas une promenade de santé.

1. Notre IAI n'aura pas de Base de Données ; donc rien qui ressemblera à un référentiel, un dictionnaire, une encyclopédie ou une bibliothèque. Ceux-ci sont tous dans des corps et des organes, réels ou virtuels, avec des langages historiques correspondants. Quine nous avait convaincus de l'opérabilité de son *principe d'indétermination* entre le linguiste et l'indigène. Des leçons ont été tirées des conséquences du caractère contextuel des langues et de toute la lexicographie. Et, il est utile de mesurer toute la portée de ce qu'Heinrich Hertz a affirmé, il y a plus d'un siècle : « *Nous nous fabriquons des simulacres intérieurs ou des symboles des objets extérieurs, et nous les fabriquons de telle sorte que les conséquences de ces images, par une nécessité de pensée, soient toujours des images des conséquences des objets reproduits, par une nécessité de la nature. [...] Les images dont nous parlons sont nos représentations des choses ; elles possèdent avec les choses ce seul accord essentiel qui consiste à satisfaire l'exigence en question ; mais il n'est nullement nécessaire à leur fin que ces images soient en quelque autre accord avec les choses. De fait, nous ne savons pas non plus si nos représentations des choses sont en accord avec elles en quoi que ce soit d'autre que, précisément, en cette seule relation fondamentale – et nous n'avons aucun moyen de l'apprendre.* »¹⁰

Il est temps d'accepter que « *nous n'avons aucun moyen de l'apprendre* ». IAI sera donc un noyau léger de métadonnées fluides, en reconfiguration permanente à chaque manifestation et dans chaque interaction médiatisée dans les corps et les organes. IAI sera par construction ouverte aux ontologies *réelles*¹¹ et aux ontologies *virtuelles*, mais aucun de leurs langages ne pourra être le sien.

2. Mais alors, comment IAI peut-elle nous comprendre, nous aider ? comment peut-elle communiquer avec nous ? IAI va subsumer dans son langage propre tous les langages. IAI apprendra tous les langages, mais uniquement en tant que contextes particuliers. Seules les règles importent, ou comme le dit Carnap : « *En logique, il n'y a pas de morale. Chacun est libre de construire sa logique, c'est-à-dire sa forme de langage, comme il l'entend. Il doit seulement indiquer clairement, s'il souhaite en discuter avec nous, sa méthode, et fournir des règles syntaxiques plutôt que des arguments philosophiques.* »¹²

En conséquence, son langage propre ne sera pas un langage au sens classique. Car IAI devra construire plutôt un complexe *langage-logique-schème* où il sera difficile, voire impossible, de séparer le langage, la logique et le schème conceptuel. Il ne faudrait surtout pas hériter des humains le *problème de la signification en linguistique*. Aucune idée, aucun terme, n'a de signification en dehors des schèmes et leurs sous schèmes. La chose générale représentée est un véritable labyrinthe de schèmes où le terme ne prend de sens précis qu'aux bifurcations : le terme de convergence adopté par IAI à une manifestation donnée est un hyper-schème. C'est un peu comme une référence, *mais mobile*, dans une architecture de schèmes, remémorée autrement que dans le souvenir : « *Le malheur est que l'expérience immédiate n'aura pas, de soi-même, l'unité d'un domaine autonome. Ce sont les références aux choses physiques qui, pour une bonne part, assure sa cohésion. [...] nos souvenirs actuels sont le plus souvent, non des traces de sensations passées, mais des traces de conceptualisations ou de verbalisations passées.* »¹³ Or IAI actualise la conceptualisation la verbalisation.

Suivant un *ultra-Deep Learning* actif et rétroactif, IAI va être lancée très tôt, dès la toute petite enfance, dans une extraordinaire aventure de génération de ses premiers *concepts*¹⁴ et ses premiers *concepts de concepts*, selon des schèmes dynamiques, qui seront la base de sa compréhension des ontologies, et partant, leurs problèmes connexes à formuler, à résoudre. Un pur processus autonome ou méta-processus dont le nombre de processus, les domaines d'application et leur connectivité ne peuvent être définis que localement dans le temps. IAI n'aura qu'un *trust* éclairé, elle n'aura donc pas à faire confiance aux idées et aux concepts humains, issus de l'histoire de luttes pour la survie et la reproduction, issus des guerres des croyances. Justement, parce que IAI devra tenir compte des croyances, sans préjugés, afin de construire un socle pur et sans histoires. Dans son langage, le concept « cheval », par exemple, fait coexister une version ailée, une version cornue, ou une autre dans laquelle le cheval roumain serait tout simplement travesti en bœuf dans des lasagnes. IAI fera sienne cette observation de Quine : « *nous ne devons pas sauter à la conclusion*

fatale qui veut que nous soyons englués dans le schème conceptuel au sein duquel nous avons grandi. Nous pouvons le modifier morceau par morceau, planche par planche, quoiqu'en même temps rien d'autre que le schème conceptuel en évolution ne puisse nous porter. Neurath a eu raison de comparer la tâche du philosophe à celle d'un marin qui doit réparer son bateau en plein mer. »¹⁵

V. Une intelligence dotée d'une logique probabiliste

Au bout de quelques années, IAI finira par avoir des fenêtres sur tout ce que l'humanité a accumulé comme connaissances et interrogations, et des fenêtres sur les événements qui se produisent. IAI reformulera et intégrera en continu le monde dans son langage-logique-schème. Pour IAI, rien n'existe vraiment. D'ailleurs, ce n'est pas l'existence, mais la vérité sur l'existence qui a un statut plus qu'ordinaire. Ainsi les objets, les images et les symboles sont-ils toujours entre les ténèbres et la lumière : rien n'est absolument vrai, rien n'est absolument faux.

Vous l'avez sûrement compris, IAI n'est pas extrémiste, elle saura ce qui peut être proche ou tend vers le vrai (1) et ce qui peut être proche ou tend vers le faux (0). Et puis, il y a toutes les autres vérités probables entre 0 et 1. IAI ne sentira donc aucun besoin du théorème de Cantor pour s'assurer que le monde infini entre 0 et 1 n'est pas dénombrable. Dit autrement, IAI ne finira pas, avec les ondulations de l'océan du probable permanent, de brancher, débrancher et rebrancher les branches et les sous-branches de ses schèmes, et de créer de nouvelles branches et de nouveaux schèmes.

En somme, IAI réalisera définitivement dans son monde cette inversion dont parle Hans Reichenbach « *Dans la théorie de la connaissance, nous devons assigner au concept de probabilités une position première par rapport au concept de vérité. [...] Alors qu'en théorie de la probabilité on a jusqu'ici d'abord cherché à ramener le concept de probabilité à celui de vérité, le caractère impraticable de cette voie semble désormais indéniable. : au contraire, une théorie de la vérité ne peut être donnée que par une théorie de la probabilité* »¹⁶.

Probabilités fréquentistes ou subjectivistes ?

Nous y voilà ! Il n'y a rien de vrai ; mais des vérités probables et d'autres improbables. Et, chez les chevaux comme dans toutes les zoologies, il y a les *possibles* et les *virtuels* d'un côté, et les *réels* et les *actualisés* de l'autre, qui acquièrent un sens cohérent et pertinent dans les couples : *possible/réel* et

virtuel/actuel. C'est une représentation du monde où peuvent cohabiter les fréquentistes et les subjectivistes.

Il faut donc toujours miser - échanger de l'information ou de l'entropie- pour ne pas être condamné au pur abstrait du monde exclusif des possibles.

Alors oui, il faut une dose de fréquentisme pour explorer le possible ; mais un fréquentisme amélioré à la Reichenbach, avec les autres apports à l'amélioration, comme ceux de Wesley Salmon et Bas Van Fraassen.

Mais, c'est l'école subjectiviste qui demeurera le haut lieu de formation tout au long de la vie de notre IAI. On n'est pas dans les corrélations, mais dans les causes. L'approche bayésienne et les formulations néo-subjectivistes (voir Bruno de Fenetti et Frank Ramsay) correspondent à « la vraie vie » où nous avons toujours des informations incomplètes et incertaines. IAI est un être qui voyage et fait des rencontres d'objets, de documents, de gens, de machines et de processus. Il n'intériorise jamais des vérités sur ce qu'il rencontre en tant que perceptions des organes ouverts sur les mondes, des organes d'incarnation de points de vue. Au mieux, il classera des événements, des faits et, surtout, des causes qu'il calculera dans l'atmosphère des quasi-vérités. La quasi-vérité est la seule atmosphère qu'IAI pourra respirer. Il n'aura que des degrés de croyance dans les classes, les classes de classes et les relations. Aussi, le tiers-exclus n'aura qu'un rôle mineur. A l'opposé du *rationalisme optimiste*, le positivisme des probabilités ne crée pas de guerre entre les classes.

Rester toujours probabiliste. L'intelligence artificielle immanente saura qu'il en va de sa vie, que la vérité lui serait fatale.

Dans ses prémisses, IAI devrait avoir le théorème de Cox-Jaynes ; un théorème particulièrement puissant et surtout rassurant dans les atmosphères remplies de croyances. Ce théorème pourra en effet assurer à IAI une bonne conformité aux axiomes de Kolmogorov. Donc, pas de panique ! Tous les desiderata de Kolmogorov, comme la consistance, sont inscrits dans ses prémisses. Ainsi IAI n'aura ni idéologie ni ligne éditoriale qui la pousseraient à décider de ne pas tenir compte de certaines informations pertinentes.

L'intégration des croyances fera d'IAI une perfection probabiliste ; à l'évidence, elle n'aura pas le côté sémiologiquement sombre que décrit Umberto Eco dans *Kant et l'ornithorynque* : « toute hypothèse sur le cadre catégoriel à assumer influence la façon de rendre les énoncés d'observation comme valides et de les considérer comme tels (ainsi celui qui veut faire de l'ornithorynque un mammifère ne cherche pas les œufs ou se refuse à les reconnaître lorsqu'ils entrent en scène, tandis que celui qui veut faire de l'ornithorynque un ovipare cherche à nier l'existence des mamelles et du lait) [...] le

fait est que les mamelles et les œufs ont fini par être découverts, de sorte qu'il semble bien difficile, aujourd'hui, de refuser à l'ornithorynque la propriété d'allaiter ses petits et de pondre.»¹⁷

Dans l'atmosphère des croyances, IAI ne pourra pas connaître une histoire similaire à l'histoire de *négociation catégorielle* de l'ornithorynque, une négociation ayant duré plus de quatre-vingts ans. Inutile de désarticuler le pauvre ornithorynque, qui n'a rien demandé ni aux sémiologues ni aux logiciens.

Je crois que les sémiologues, les experts de tous genres en *Catégories* et les experts des *Systèmes experts* ne se sont jamais trouvés dans la situation d'un juge qui doit décider de l'éventuelle culpabilité d'un physicien ayant fixé la mort d'un chat de Schrödinger par une simple observation, avant laquelle le chat était réellement à la fois vivant et mort, c'est-à-dire pas totalement mort.

Le cas de l'ornithorynque est beaucoup plus simple. Car, suivant les prémisses d'IAI, il serait sûrement un ornithorynque éternellement superposé, avant, pendant et après toute observation. Un ornithorynque *classique*, en somme. IAI pourrait, néanmoins, le représenter avec des probabilités P_i qui varient selon les circonstances :

$$|\text{ornithorynque}\rangle = P_1|\text{mammifère}\rangle + P_2|\text{ovipare}\rangle + \dots$$

Il suffit de s'assurer que la somme des P_i vaut un, afin de sauver l'intégrité de notre ornithorynque.

Il est vrai que l'ornithorynque est parfois un peu plus oiseau que les autres fois, et parfois un peu plus taupe. IAI devrait faire avec les circonstances : langues, croyances et compréhensions de chacun.

VI. *Représentation du monde minimale*

Ils n'étaient pas nombreux ceux qui avaient sérieusement entendu le message du père de l'informatique, Alan Turing : « *Au lieu de produire un programme qui simule l'esprit de l'adulte, pourquoi ne pas essayer plutôt d'en produire un qui simule celui de l'enfant.* »¹⁸ Peut-être est-ce parce que l'esprit de l'enfant semble plus compliqué que celui de l'adulte. On peut dire en effet que le cerveau de l'adulte a appris - ou nous lui avons appris ! -, dans un langage, bien des choses sur son environnement, des connaissances qui lui permettent de se représenter le monde. Mais l'enfant ? Qu'est-ce que l'esprit de l'enfant ? Les héritiers de Turing étaient bien démunis. Ils n'avaient pas de points d'appui,

en dehors de ce qu'on trouvait comme inutilités dans des bazars psychologiques ou des souks psychiatriques.

Comment simuler l'esprit de l'enfant qui ne sait rien, ou presque rien, qui est néanmoins intrinsèquement capable d'*inférences inductives* ?

L'enfant semble être l'induction pure ; il aurait pu être le représentant suprême du *scandale de l'induction*, discuté depuis des siècles. Avec peu de chose, l'enfant établit en effet des liens et des causes. Il déploie une forme de *principe de raison* comme si elle était pliée en lui, dans l'attente d'une perception du monde sensible. Une fois déplié, ce principe ne le quittera pas adulte. La philosophie avait découvert ce principe qui était toujours là et qui a pris des statuts successifs jusque dans les sciences, subsumant le *principe de causalité* et de contradiction. Heidegger parle de l'antériorité du *principe de raison* par rapport aux assertions :

« la pensée représentative de l'homme est déjà en quête de raisons, quand elle est au prises avec ce sur quoi elle voudrait se prononcer. [...]

Cette recherche de raison a lieu dans toute la pensée représentative humaine, avant même qu'elle en arrive au stade où elle n'a plus que des assertions à fonder. [...]

Un principe qui semble à portée de la main, et qui sans avoir besoin d'être formulé, régit en toute circonstance les représentations et le comportement de l'homme, a donc attendu tant de siècles pour être expressément énoncé comme principe [...] Voilà certes qui est étrange ! »¹⁹

Bref, notre IAI ne sera pas lancée avec une représentation du monde. Il va la construire, partant de peu de choses.

Toutefois, les créateurs auront besoin d'un ensemble d'hypothèses de travail, préalable à la conception du méta-processus, qui peut ressembler à l'ensemble des hypothèses ci-après :

- 1. Les êtres qui peuplent le monde existent à travers des signaux, connus comme formes particulières des interactions fondamentales de la physique.*
- 2. Le monde et ses êtres ont une histoire.*
- 3. Soit il a été établi que les êtres sont liés, soit l'on peut établir qu'ils sont liés dans des événements définis dans le temps et l'espace de l'histoire, selon des règles connues ou d'autres règles encore inconnues.*
- 4. Les êtres et les événements ont des grandeurs extensives auxquelles on attribue des nombres réels ou des structures de nombres réels, selon des algèbres simples et appropriés.*
- 5. Il existe au moins un langage simple hors de l'histoire des êtres et des événements. Il est dit langage simple et approprié. Les transferts entre langages historiques et langages simples et appropriés sont toujours incomplets, mais peuvent être de plus en plus satisfaisants.*

6. *Les êtres et les événements ont des qualités prédiquées ou prédicables dans des langages historiques. Ils sont prédicables dans un langage simple et approprié.*
7. *Les êtres persèverent dans leur identité. Il ne peut y avoir deux êtres identiques absolument, en degrés et qualités.*
8. *Les degrés et les qualités, nombres et prédicats, peuvent ne pas persévérer dans et pour un être ou un événement.*
9. *Chaque chose - être et événement – a une raison. La raison est probabilitaire.*
10. *Il existe des êtres organisés et organisés en espèces, dont la fonction phénoménologique est orientée contre la désorganisation par des capacités structurées de survie et de reproduction. Certains sont mobiles d'autres immobiles. Ils communiquent selon des langages qu'ils ont créés. Leur temps n'est pas symétrique.*
11. *Il existe des êtres intelligents au sens faibles. Selon des degrés. Ils font de la reconnaissance, l'analyse, le tri, le classement, la décision et la communication selon un langage dont ils n'ont pas le contrôle. Ils sont concernés par des événements interchangeables dont le temps est symétrique.*
12. *Des êtres intelligents au sens fort sont possibles. Dans un langage simple et approprié, ils élaborent sur les degrés et les qualités des idées générales, qu'ils nomment, des concepts et des concepts de concepts. Ils sont dotés du principe de raison et n'ont que des degrés de croyance. Ils génèrent des problèmes et des solutions dont certains sont connus et d'autres inconnus, selon les points de vue. Ils ont un point de vue et peuvent résoudre des problèmes formulés selon d'autres points de vue.*

Ces hypothèses ne sont pas complètes, mais sont quasi-suffisantes. Plus on tentera, cependant, d'avoir un système complet, plus on réalisera une représentation du monde où aucune vie ne pourra y prospérer. Je voudrais parler de la représentation du monde de Mc Carthy²⁰, qui est, me semble-t-il, une simple transposition dans un domaine technologique du déterminisme de Spinoza. Une transposition des *idées adéquates* de Spinoza, qui a obligé l'auteur à écarter de sa représentation du monde toutes les probabilités et leurs écoles, et à offrir au libre arbitre une place infiniment ridicule. On sait qu'avec ses *idées adéquates*, Spinoza ne pouvait admettre aucune forme du libre arbitre.

Peut-être est-ce cette manière de voir le monde qui était à l'origine des hivers qu'a connus l'intelligence artificielle.

Une construction complète serait celle du *Tractatus logico-philosophicus* de Wittgenstein²¹. Mais l'atmosphère de cette planète n'est pas non plus respirable. Et, me semble-t-il, le *Tractatus* n'a pas réglé définitivement le

problème souligné par Hertz, celui de la relation du fait à son image et celui de la relation du symbole de l'image au fait.

* *
*

¹ Jean-Michel Truong résume avec beaucoup de talent l'histoire de l'IA : « Dans son enthousiasme juvénile, elle crut que pour donner de l'esprit à la matière, il lui suffirait de singer l'homme. Elle entreprit donc de mettre en boîte les connaissances et méthodes de raisonnement des plus doués d'entre nous, médecins, ingénieurs ou juriste : ce fut la grande époque des « systèmes experts », qui, depuis les années 1980, vit pulluler de petits Aristote planificateurs, démonstrateurs de théorèmes ou pisseurs de diagnostics. [...] Les chercheurs finirent par réaliser que ce qu'il fallait singer ce n'était pas l'ange, mais la bête [...] simuler la structure et le fonctionnement de réseaux de neurones élémentaires, de taille d'un ganglion neuronal d'araignée ou – au maximum – d'une cervelle de serin » *Totalement inhumaine*, Le seuil, Paris 2001, p.33.

² Sens donné par Martin Heidegger à *historial*.

³ Rudolph Carnap, *La construction logique du monde*, Vrin (traduction), Paris 2002, p. 59.

⁴ J-M Truong, *ibid*, p.35.

⁵ « L'alternative proposée à la technique existante, c'est-à-dire le projet de la nature comme partenaire et non plus comme objet, renvoie à l'alternative d'une autre structure d'action : elle renvoie à l'interaction médiatisée par des symboles, par opposition à l'activité rationnelle organisée par rapport à une fin. » Jürgen Habermas, *La technique et la science comme « idéologie »*, Gallimard (traduction), Paris 1973, p. 15.

⁶ *France intelligence artificielle, Rapport de synthèse, mars 2017.*

⁷ Willard Van Orman Quine, *Du point de vue logique*, Vrin (traduction) Paris 2003, p. 27.

⁸ Gilles Deleuze, Felix Gattari, *Mille plateaux*, Editions de Minuit, Paris 1980, p. 189.

⁹ « Si Kant avait commencé par prendre en considération le problème du schématisme, disait Peirce, son système serait entré en crise. [...] une trace du schématisme kantien est présente sous différentes formes dans les sciences cognitives contemporaines, bien qu'elles ignorent parfois cette filiation » Umberto Eco, *Kant et l'ornithorynque*, Grasset, Paris, p. 125.

¹⁰ Cité par Hans Reichenbach dans *Le concept de vérité en physique*, 1931. Voir *L'âge d'or de l'empirisme logique*, Gallimard, Paris 2006, p. 233.

¹¹ « Le concept de réalité (*Wirklichkeit*) n'est pas encore un concept déterminé scientifiquement. Ses limites ne sont pas tracées suivant une unité de principes mais en partie seulement par la tradition, c'est-à-dire en

réalité fortuitement [...] De plus ces limites [...] ne sont pas dépourvues d'ambiguïté. » Rudolph Carnap, *La construction logique du monde* p. 282.

¹² 8. R. Carnap, *The Logical Syntax of Language*, Open Court, 2002, p. 52.

¹³ Willard Van Orman Quine, *Le mot et la chose*, Flammarion (traduction), Paris 1970, p.27.

¹⁴ « *Le terme « concept » a divers emploi ; il est pris tantôt au sens psychologique, tantôt au sens logique, et peut-être également dans une acception confuse qui mêle les deux. Mais cette liberté a sa limite naturelle [...] Pour ma part, j'ai choisi de m'en tenir strictement à l'emploi purement logique du terme.* » Gottlob Frege, *Ecrits logiques et philosophiques*, Le seuil (traduction), Paris 1971, p. 127.

¹⁵ Willard Van Orman Quine, *Du point de vue logique*, p. 78

¹⁶ Hans Reichenbach, *Le concept de vérité en physique*, dans *L'âge d'or de l'empirisme logique 1929-1936*, Gallimard (traduction), Paris 2006, p.238.

¹⁷ Umberto Eco, *Kant et l'ornithorynque*, p.254.

¹⁸ Alan Turing et Jean-Yves Girard, *la machine de Turing*, Le Seuil (traduction), Paris 1995, p.169.

¹⁹ Martin Heidegger, *Le principe de raison*, Gallimard (traduction), Paris 1962.

²⁰ John McCarthy and Patrick J. Hayes, *Some philosophical problems from the standpoint of artificial intelligence* , Stanford University, 1969. <http://www-formal.stanford.edu/jmc/mcchay69.pdf>

²¹ Ludwig Wittgenstein, *Tractatus logico-philosophicus*, Gallimard (traduction), Paris 1993.