

Au sujet de l'histoire de l'intelligence artificielle (ia)

Stefan Padberg

Étant donné que cela plaisait peu au roi que son fils,
abandonnant les routes contrôlées, courût çà et là
à travers champ pour se faire lui-même un jugement
sur le monde, il lui offrit un carrosse et un cheval.
« Tu n'as plus besoin de marcher », furent ses mots.
« Désormais tu ne le devras plus », en fut le sens.
« Désormais tu ne le sauras même plus », le résultat¹

L'histoire de ce qu'on appelle l'intelligence artificielle (ia) est une histoire pleine de méprises et de confusions. Mais avant tout c'est l'histoire d'une mésintelligence de soi de l'être humain car le fossé entre ce qui se passe, réellement au niveau technique et ce que les êtres humains *croient* qu'il se passe est large et il n'a cessé de s'élargir de plus en plus au cours de l'évolution. Même les résultats réels de cette technologie se distinguent nettement de son efficacité rêvée (ou de ses craintes). Dans l'exploration suivante, cette évolution sera remémorée à l'appui de quelques stations auxquelles ce processus se laisse illustrer de manière symptomatique.

Alain Turing

Pour le développement du concept « d'intelligence artificielle », la connaissance d'Alan Turing et de ce qu'il a lui-même désigné comme un « test », est extrêmement profitable, mais révèle déjà des tendances essentielles, qui nous occupent encore aujourd'hui. Alan Turing, né à Londres en 1912, était un mathématicien anglais très doué. Il fut au plus célèbre pour avoir apporté la contribution décisive permettant de déchiffrer les informations radio allemandes transmises par le système de codage *enigma* au début de la seconde Guerre mondiale. Mais en informatique son nom est avant tout rattaché au concept de « machine de Turing » qu'il introduisit en 1936, dans l'informatique naissante. Il s'agissait alors d'un modèle de calculateur abstrait, avec lequel on pouvait étudier des algorithmes, avant même qu'il existât encore de véritable calculateur.

Alan Turing était un mathématicien dans l'âme. À 16 ans environ il lut le travail d'Albert Einstein au sujet de la théorie de la relativité et il fut, à l'époque, l'un de ces rares êtres humains capables de suivre complètement de bout en bout à l'époque les réflexions mathématiques complexes d'Einstein [Pour les calculs mathématiques indispensables à sa théorie de la relativité, Einstein fut par ailleurs redevable à son épouse, une mathématicienne exceptionnelle qui avait renoncé à sa passion pour s'occuper de leurs enfants, *ndt*]. Sa querelle avec Ludwig Wittgenstein est aussi bien connue, en 1939 à Cambridge, quant à savoir si la mathématique pût mettre à jour des vérités absolues. Alors que celui-ci tenait les vérités mathématiques pour sur-évaluées, Alan Turing insistait sur le fait qu'avec l'aide du formalisme mathématique des vérités absolues pouvaient être découvertes. En 1953, il développa son premier programme de jeu d'échec pour lequel il n'existait encore aucun *hardware* à l'époque, de sorte que chaque coup devait être calculé « à la main », selon l'algorithme développé par lui. En 1954 Alan Turing mit fin volontairement à sa vie. Il avait été condamné deux ans auparavant à la castration chimique, à cause de son homosexualité ce qui entraîne comme effet secondaire une grave dépression.

Le test de Turing

Dans un travail de l'année 1950² il se préoccupa de la question de savoir si des machines peuvent penser et donc de la comparabilité des productions du calcul sur ordinateur avec l'intelligence (humaine), bien qu'à l'époque on n'était pas du tout en situation de construire de telles machines. Étant donné qu'on ne pouvait pas définir les concepts de « machine » et de « penser » d'une manière universellement valable, il proposa un « jeu d'imitation » : une machine peut-elle imiter un être humain de manière convaincante au point qu'un autre soit incapable d'en reconnaître la différence ?

Au cours de ce test, un interrogateur humain, devant un clavier et un écran, sans contact visuel ou auditif, mène un entretien avec deux partenaires inconnus de lui. L'un des deux est un être humain, l'autre une machine. L'interrogateur, à la suite de l'interrogation intensive n'ayant pas pu pas dire clairement lequel de ses deux partenaires est la machine à l'issue du test, celle-ci a donc passé le test et on imputa donc à la machine une même capacité de penser.

¹ Günter Anders, *Die Antiquiertheit des Menschen [La désuétude de l'être humain]*, Munich 2002, p.97

² Alan Turing, *Computing Machinery and intelligence*. *Mind* 49, pp.433-460, web : <https://academic.oup.com/mind/article/LIX/236/433/986238>

L'effet Turing

Dans ce dispositif (*setting*, en anglais dans le texte), qui entra par la suite dans l'histoire, on n'avait donc pas testé si la machine pouvait penser mais plutôt seulement si des êtres humains se laissaient abuser par des machines, si des machines pouvaient imiter des êtres humains.³ Alan Turing tint pour possible qu'à la fin du (20^{ème}) siècle, une machine pût « passer avec succès » son test. Il tenait cela de manière prépondérante comme un pur problème de programmation, et bien moins comme un problème de *hardware*. Son estimation est presque déjà bouleversante, à savoir que le « contenu d'information » (« *storage capacity* ») du cerveau humain se situât entre 10^{10} et 10^{15} *bit*, mais que pour une communication réduite, sans participation visuelle, telle qu'elle était supposée dans son *setting*, 10^7 bits seraient seulement nécessaires. Quand on sait que les capacités de calcul mises en jeu aujourd'hui sont plus grandes à beaucoup d'égards, sans que l'on eût créé le moins du monde ce qui pût passer le test de Turing avec succès, on ne peut que s'étonner de ces réflexions. Au cours de notre exploration, nous buterons sans cesse sur ce modèle de sous-estimation de soi et de surestimation de la machine. C'est pourquoi je voudrais proposer la caractérisation « d'effet Turing » pour désigner ce phénomène.

Intelligence abstraite

La deuxième chose qui frappe, c'est que Alan Turing, sans le dire ici, détache le concept d'intelligence de son fondement biologique. Dans ses réflexions, il n'existe que l'intelligence mathématique, ou selon le cas tous les phénomènes se laissent transposer en elle. Il classe même dans cette catégorie des processus d'apprentissage d'un petit enfant et il va jusqu'à imaginer, à la fin de son article de construire au mieux une machine à apprendre que l'on pût ainsi entraîner de la même façon qu'on instruit un petit enfant :

« Dans la tentative d'imiter un esprit humain adulte nous sommes tenus de beaucoup réfléchir sur le processus qui l'a amené à l'état dans lequel il se trouve. Nous pouvons en remarquer trois composantes.

- a) l'état du début de l'esprit, par exemple à la naissance,*
- b) la formation à laquelle il fut soumis,*
- c) d'autres expériences qui ne sont pas à caractériser comme une formation, et auxquelles il fut soumis.*

Au lieu d'essayer de développer un programme de simulation de l'esprit adulte, pourquoi ne pas faire de préférence un programme qui simule celui de l'enfant ? Si celui-ci était ensuite soumis à une formation correspondante, il en résulterait le cerveau de l'adulte. L'enfant est probablement quelque chose comme un carnet de notes comme on en achète chez le papetier. Plutôt moins de mécanisme et de nombreuses pages vides. (Mécanisme et écriture signifient la même chose à partir de notre vision). Notre espoir c'est qu'il y ait si peu de mécanisme dans le cerveau de l'enfant, de sorte qu'il puisse être facilement programmé quelque peu dans son genre. Pour la dépense en travail dans l'éducation, on peut admettre en première approche qu'il correspond à peu près à celui engagé pour l'enfant humain. »⁴

Au plan biographique on peut admettre que Turing vécut ainsi totalement dans le monde des abstractions mathématiques. Celles-ci sont objectivement données à l'être humain et sont identiques pour tous les êtres humains. Ainsi n'a-t-il jamais remarqué que le développement de l'intelligence humaine est avant tout lié à la confrontation de l'esprit humain individuel avec la Terre. Or ce **concept abstrait d'intelligence**, qui n'est quasiment jamais associé à une biographie humaine ou à une production cognitive, émerge toujours sans cesse à nouveau au cours de l'histoire du développement de l'ia.

La conférence de Dartmouth

À ma connaissance, Alan Turing n'avait pas encore connu ni utilisé le concept d'ia (*artificiel intelligence* en anglais). Celui-ci surgit la première fois dans le champ avancé d'une conférence qui fut tenue en 1956 au *Dartmouth College* sur la côte est des USA. Celle-ci fut préparée et réalisée par de jeunes scientifiques, John McCarthy et Marvin Minsky. Dans la feuille d'annonce de la *Rockefeller foundation*, il était dit entre autres :

« Nous proposons de mener à bonne fin un séminaire de deux mois, au cours de l'été 1956, au sujet de l'intelligence artificielle avec 10 participants au Dartmouth College. L'étude doit partir de l'hypothèse que tous les aspects de l'apprentissage et d'autres caractéristiques de l'intelligence peuvent être fondamentalement décrits aussi précisément qu'une machine puisse en être construite pour la simulation de ces processus. On va tenter de découvrir comment des machines peuvent être amenées à utiliser un langage, entreprendre des abstractions, développer des concepts, des problèmes du genre de ceux

³ Nous rencontrons ce genre de test simple de Turing sur *internet* lorsque nous devons résoudre une énigme **CAPTCHA**, un acronyme qui signifie : « *Completely Automated Public Turing test to tell Computers and Humans Apart* ».

⁴ *op. cit.* §7 *Learning Machines*. [traduction en allemand de S.P.]

qu'actuellement l'être humain se réserve de résoudre et d'améliorer lui-même. Nous croyons que des progrès peuvent être envisagés et ciblés dans l'un ou l'autre de ces champs de problèmes, si un groupe soigneusement composé de scientifiques y travaille tout un été durant. »⁵

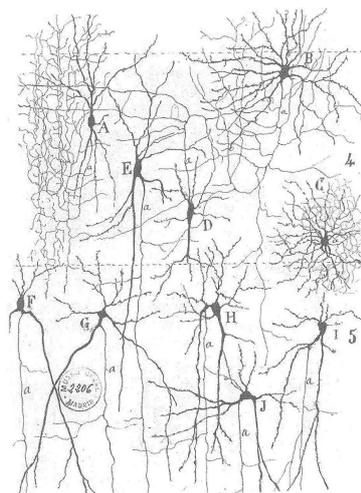
Comme sous-thèmes, dont la conférence devait se préoccuper, les domaines suivants étaient désignés dans la proposition du projet :

1. Ordinateur automatique ;
2. Comment programmer un ordinateur pour utiliser un langage ;
3. Réseaux neuronaux ;
4. Réflexions théoriques sur l'extension d'une opération de calcul ;
5. Auto-amélioration ;
6. Abstractions ;
7. Contingence et créativité.

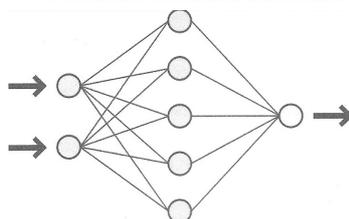
Le *Faszinosum*^(a) de l'ia

La conférence ne dura pas l'été durant deux mois, mais seulement un mois et ce ne fut rien d'autre qu'une séance relâchée de *brainstorming*^(b) entre ceux des jeunes scientifiques qui étaient intéressés à ce sujet. Elle est pourtant considérée en informatique comme la date de naissance de la recherche sur l'ia.

Bien entendu il n'y avait aucun *hardware* à l'époque avec lequel de telles *visions* eussent pu se voir fondées, et la conférence elle-même n'avait aucun concept utilisable et transposable pour un résultat. Pourtant le concept^(c) naquit ici. L'aura de fascination qui entoure depuis ce concept, déploya dès lors toute son influence à partir de cet instant. J'appelle cela l'*ia-faszinosum*^(a)



Neuronales Netz (Quelle Wikipedia)



Réseaux neuronaux artificiels

Un concept qui venait de la biologie fit son apparition à cette conférence qui doit être éclairé d'un peu plus près : celui de réseaux neuronaux. La recherche en physiologie cérébrale aux 19^{ème} et 20^{ème} siècles fournit peu à peu la conviction que le cerveau était, pour l'essentiel, un réseau hautement complexe de neurones par lesquels sont conduites des impulsions électriques (potentiels d'action). Ceci inspira le physiologiste américain Warren S. McCulloch et le logicien américain Walter Pitts, pour projeter un ordinateur qui était censé avoir une structure copiée à partir d'un réseau neuronal (cellules McCulloch-Pitts). Ils prouvèrent qu'avec de tels automates que toute fonction logique ou arithmétique pouvait se laisser calculer.

Réseau neuronal

Schématisation du réseau neuronal,

En même temps un modèle de réseau neuronal artificiel (source wikipedia)
Un neurone peut être relié à sa sortie avec plusieurs. Plusieurs neurones
Peuvent « finir » du côté de leur sortie dans un seul neurone.

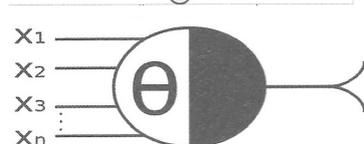


Diagramme simplifié d'une cellule McCulloch-Pitts (source wikipedia)

X_1, \dots, X_n : connexions d'entrée du signal ; θ : valeur seuil ;
tâche de l'algorithme : si la somme des signaux d'entrée est supérieure ou égale à la valeur seuil , c'est un « 1 » qui sort sinon c'est un « 0 ».

En 1949, le psychologue Donald A. Hebb, publia ses résultats de recherche sur la plasticité synaptique. Il démontra que les synapses, et donc les liaisons entre les cellules nerveuses, se multiplient ou bien se

⁵ J. McCarthy/M. L. Minsky/N. Rochester/C. E. Shannon/, *A Proposal For The Dartmouth Summer Research Project On Artificial Intelligence*, Aug. 31, 1955, *Webs* : <https://web.archive.org/web/20080930164306/http://www-formal.stanford.edu/jmc/history/dartmouth.html>

réduisent à chaque fois selon la fréquence avec laquelle elles sont sollicitées.⁶ Avec cela, on crut avoir découvert les correspondances physiologiques des processus de l'apprentissage. Cela inspira de surcroît la mise au point de Réseaux Neuronaux Artificiels, dans lesquels la valeur seuil θ des neurones artificiels isolés se laisse modifier en dépendance de la réussite du calcul. Avec cela des programmes seraient possibles qui s'optimiseraient eux-mêmes (ceux dont on dit que ce sont des algorithmes « capables d'apprendre »). On espérait ainsi enfin découvrir une architecture de calcul avec laquelle on pourrait imiter des productions de l'intelligence humaine.

Image de soi réduite de manière matérialiste comme inspiration pour des découvertes

C'est bien entendu une faute de penser que de croire être en mesure de conclure que les résultats d'un processus d'apprentissage fussent directement identiques à des modifications au niveau des synapses neuronales et donc que les « informations », qui dépendent de ces processus d'apprentissages, eussent été « sauvées » [au sens d'emmagasinées, ici *ndt*] dans ces structures biologiques.

Cette conclusion fautive, si largement répandue, est encore et toujours due à la vision du monde réductionniste et matérialiste qui domine. En tant que neuro-biologiste profane je présumerais ici plutôt, pour ma part, une corrélation avec des processus d'apprentissage « qui vivent longtemps », par exemple la formation de modèles de comportement se déroulant inconsciemment, comme ceux nécessaires à conduire une voiture, par exemple. De même la totalité de notre organisation sensorielle est imprégnée de telles réflexes qui se déroulent inconsciemment et que nous avons appris dans la tendre enfance, et donc pour ainsi dire modelés à fond. À l'issue d'un processus d'apprentissage et d'imprégnation cognitifs plus longs, de telles structures biologiques existent ensuite, que nous pouvons percevoir et engager à l'horizon de notre conscience comme des facultés, sans devoir pour autant y réfléchir de nouveau.

Malgré cela, à l'aide de tels réseaux de neurones artificiels, on croyait pouvoir édifier réellement en le copiant le cerveau et ses productions. C'est un curieux paradoxe que sans cesse dans l'histoire des temps modernes des découvertes ne cessent d'émerger : l'image de soi de l'être humain, en tant qu'il se la représente comme un mécanisme machinal inspire la réelle découverte de machines.

[l'adjectif « machinal » ici est pris au sens français originel de « ce qui appartient à la machine », point 1 & 2 de la rubrique du Littré à ce terme : tome 4, p.3620, *ndt*]

« Ordinateur neuronal »

Marvin Minsky, mentionné déjà lors de la conférence de *Dartmouth*, s'était toujours illustré dans ce domaine par ses recherches. En 1951, déjà il avait soutenu sa thèse en développant un ordinateur neuronal, le *snark*, qui était capable de fixer automatiquement sa valeur seuil de potentiel. Le premier « ordinateur neuronal » réussi (*Mark / Perceptron*) fut développé dans les années 1957-1958 par Franz Rosenblatt, Charles Wightman et collaborateurs et mis en œuvre pour des problèmes de reconnaissance de modèles. Avec un grand détecteur d'image (20 x 20 pixel), il pouvait reconnaître de simples chiffres et fonctionnait avec l'aide de 512 potentiomètres motorisés, chacun d'eux ayant des seuils de potentiels variables.

Le premier hiver de l'AI

Reconnaître un paragon d'une programmation sérielle classique est très complexe et coûteux en calculs, ce qu'on ne pouvait pas résoudre avec le *hardware* de l'époque. Dans cette mesure, l'amorce engagée avec les ordinateurs neuronaux, apparaissait être une voie très élégante et très prometteuse. Dans l'analyse mathématique du *perceptron* que mena Marvin Minsky, en 1969, avec Seymour Papert, il fut pourtant constaté que cette amorce technique à partir de réflexions de base, ne pouvait mener qu'à des résultats partiels et que c'était véritablement un cul-de-sac technologique. Or là-dessus, un afflux d'argent important avait été engagé, présumé aussi rapporté beaucoup par la suite et pour la première fois, on mit donc ce domaine de recherche à l'arrêt pour un temps plus long.

Ceci n'est pas un événement inhabituel dans le paysage de la recherche aux USA. C'est un pays où cette recherche dépend très fortement de sources d'argent extérieures, il faut donc y raconter de belles histoires pour acquérir cet argent avec succès et cela est soumis en outre aux tendances de la mode (*Modetrends*), qui résultent souvent des débats publics. Ceux-ci sont fortement imprégnés de curiosité, de joies dans

⁶ Donald O. Hebb, *The organisation of behavior* Erlbaum Books, Mahawah, N.J. 2002 (reproduction de l'édition de 1949). La première règle qui y est formulée a la teneur suivante : « Lorsque qu'un axone de la cellule A reste assez près pour exciter une cellule B et de manière répétée ou persistante engendrer le potentiel d'action de la cellule B, alors certains processus de croissance ou de modifications métaboliques ont lieu dans l'une ou dans les deux cellules, de sorte que l'efficacité de la cellule A, en est plus grande comme l'une des cellules excitantes sur la cellule B. »

l'expérimentation, de pragmatisme et de la disposition à prendre des risques. Même un film bien fait qui est préparé à partir d'un sujet de recherche captivant déterminé, peut rehausser considérablement chez les *sponsors* potentiels la joie d'apporter des fonds.

Le militaire suit purement et simplement avec une opiniâtreté relative son propre agenda de recherches, mais en définitive, il reste dépendant dans la fixation de ses objectifs, à son tour aussi de tout ce qui circule comme idées dans la communauté de recherche. Or justement, cette marche-arrière de presque 15 ans, parle en faveur du fait que les programmes de recherche aux USA sont nettement moins stratégiquement planifiés qu'il est communément admis qu'ils le sont.

Marvin Minsky resta cependant optimiste. Quoiqu'il n'y avait pas la moindre occasion pour cela du côté de l'évolution technique du *hardware*, il crut même pouvoir prédire qu'il serait bientôt possible de « programmer des émotions dans une machine ». En 1970, il déclara que dans trois à huit ans, il y aurait des machines dotées de l'intelligence moyenne d'un être humain, qui liraient Shakespeare et feraient l'entretien des autos. Il a été ici nettement subjugué par un effet Turing comme il est caractérisé plus haut.

L'interface être humain-machine

Dans l'intervalle, d'autres problèmes de recherche furent explorés. Au milieu des années 60, Josef Weizenbaum commença à se préoccuper de la manière dont un être humain peut communiquer avec une machine par un clavier dans un langage humain. Il rédigea pour cela un programme nommé ELIZA. Dans une version simple, il utilisa un thésaurus et donc un dictionnaire structuré. La phrase qui y est donnée est recherchée pour le cas où elle renferme un mot qui existe dans le thésaurus, ensuite, en partant de ce mot, on cherche des synonymes, le plus souvent cependant cherchés selon des concepts génériques. Le programme renferme un recueil de phrases au sujet de divers domaines de thèmes. Le programme pouvait simuler divers partenaires dialogiques. La simulation d'un psychothérapeute, qui travaillait selon la méthode de l'entretien décentré sur le patient de Carl Roger, connut un succès très important. On peut recourir à un exemple de paires de questions-réponses telles que celles-ci :

Utilisateur : « J'ai un problème avec mon **père** . »
ELIZA : « Dites-moi en plus sur votre **famille** ! »

Dans ce cas le programme a transposé l'ordre de *père* en son concept générique de *famille* et il a été recherché une phrase au sujet du complexe de thèmes *famille*. Cette manière simple de fonctionner révèle la limitation du programme, que l'utilisateur ne comprend en aucun cas. Ainsi, le principe du mot-clef peut aussi mener au semblant de « dialogue » suivant :

Utilisateur : « La guerre est le **père** de toutes choses . »
ELIZA : « Dites-moi en plus sur votre **famille** ! »

Caractéristique, pour ce genre de conduite dialogique, c'est la répétition de ce qui est dit ou la reformulation d'une phrase énoncée dans une interrogation :

Utilisateur : « Hello ! Mon ami m'a dit que je devais venir ici. »
ELIZA : « Ton ami t'a dit que tu dois venir ici ? . »
Utilisateur : « Il affirme que je suis souvent dépressif. »
ELIZA : « Je suis désolé si tu es souvent dépressif. »
Utilisateur : « C'est cela. Je ne suis pas heureux . »
ELIZA : « Penses-tu que venir ici peut t'aider à ne plus être malheureux ? »

Quand le programme ne trouve pas de mot avec lequel commencer quelque chose, il répond par des phrases générales évasives, du genre : « Je n'ai pas encore compris cela ; pouvez-vous m'expliquer cela. » Ou bien : « Je ne comprends rien à cela, parlons d'un autre sujet. »

L'effet ELIZA

Les personnes de l'essai furent quant à elles bel et bien convaincues que le « partenaire dialogique » dans les expérimentations, apportait une compréhension effective à leurs problèmes. Elles commencèrent à attribuer à la machine des sentiments et une compréhension. Même lorsqu'elles furent confrontées au fait concret qu'elles avaient « parlé » avec le programme de l'ordinateur et que sur la base de quelques règles simples et assurément sans « intelligence » ni « compréhension », ni même « capacité d'identification » ou autres, celui-ci avait transformé des affirmations données en problèmes, elles refusèrent souvent de l'accepter.

Au moment où Josef Weizenbaum, un jour, surprit sa secrétaire en train de « converser » avec le programme sur ses problèmes les plus intimes, laquelle lui fit quitter aussitôt la pièce au nom de la « protection de la sphère privée », il fut très profondément estomaqué de constater de telles réactions à son programme. Même des praticiens psychothérapeutes crurent sérieusement pouvoir parvenir avec cela à une forme automatisée de psychothérapie. Cette projection de qualités et de facultés humaines dans la machine fut dorénavant désignée comme « l'effet-Eliza ». Or cet effet ne cesse de surgir et il est même souvent consciemment amorcé par une partie de la scène-ia tandis que celle-ci tente de faire ressembler ses machines le plus possible aux humains. Le problème là-dedans, c'est que l'on se met alors à vivre dans un semblant de réalité et que l'on accorde aux machines quelque chose comme un « crédit de confiance » qui empêche ensuite à tout un chacun de faire face aux activités de la machine de manière critique lorsque cela est indispensable. — Joseph Weizenbaum devint donc sur la base de ces expériences et de leur constat, le premier informaticien critique au monde.

Réseaux Neuronaux Artificiels (RNA) avec retour sur erreurs

Dans le premier hiver-ia, qui dura de la fin des années 1960 jusqu'au milieu des années 1980, de nouvelles topologies de réseaux furent explorées. Une amorce très prometteuse consistait à renvoyer l'erreur en retour au RNA et donc, la déviation entre l'édition calculée et celle souhaitée. Avec certains algorithmes affinés, ce retour de communication fut à l'occasion utilisé pour améliorer les valeurs seuil des neurones artificiels isolés. Ainsi put-on optimiser un RNA.

Le mathématicien américain et spécialiste de neurologie, John Hopkins, put finalement se représenter en 1985 la solution du problème du représentant de commerce. Il s'agissait là du calcul de l'itinéraire optimal d'une tournée de représentation commerciale. Ainsi fut-il en mesure de démontrer que les Réseaux de Neurones Artificiels n'étaient pas du badinage, mais convenaient foncièrement à la résolution de problèmes pratiques. Par la suite de plus grosses sommes d'argent de recherche se remirent à affluer dans cette technologie.

Avec l'aide de cette technique, des problèmes de déchiffrages et de décryptages⁷ purent être automatisés. À partir du milieu des années 1980, la distribution du courrier fut presque totalement automatisée en Allemagne, car l'identification du code postal put être automatisée. Ce fut la raison pour laquelle la poste modifia son format d'adresse. Si jusque-là on écrivait les éléments de l'adresse dans la succession suivante : Nom — code postal — ville — rue, on dut dès lors les ordonner dans le sens : Nom — rue — code postal — ville ; De ce fait fut garanti que le code postal se trouvait toujours en bas à gauche sur l'enveloppe. Sous cette condition, le modèle de reconnaissance fut complètement automatisé, le code postal déchiffré et la lettre put se retrouver dans le bon casier.

À ma connaissance, ceci fut la première utilisation qui intervint réellement en profondeur dans la vie du travail, car le nombre des employés au tri postal put être réduit de manière drastique. Le nombre des emplois perdus dut s'élever en moyenne à cinq chiffres.

À cette époque il y eut aussi les premières tentatives de programmation d'automobiles autonomes. Mais celle-ci ne pouvait pas encore être transposée car on ne disposait pas encore des capacités de calcul nécessaires. L'auteur se souvient encore très bien qu'une équipe dans une grosse firme, pour laquelle le positionnement de tâche était familier, mettait régulièrement en rade le calculateur central, lorsque le programme de simulation de conduite était lancé. Les essais furent finalement arrêtés. Mais il était bien établi que les réflexions correspondantes pour l'automobile autonome se trouvaient déjà à l'horizon des développeurs techniques au milieu des années 1980 [le milieu des années 80 marque aussi dans les laboratoires de biochimie, la conduite programmées sur ordinateur (et l'apparition pour ce faire déjà de la « souris » de *Microsoft* : 1987, pour ma part) de l'appareillage des séparations de protéines et de l'analyse des acides animés en routine, avant cela tout fonctionnait encore avec des vannes « électromécaniques » minuteriers programmées « à la main », avec des *ndt*].

« Deep blue » vainc Garri Kasparov

Une équipe de chez IBM tenta de contourner la limitation causée par la capacité de prestation insuffisante du *hardware* en édifiant un « super calculateur ». L'objectif était de construire un ordinateur de jeu d'échec qui était censé battre les maîtres mondiaux professionnels du jeu d'échec. Le super-calculateur ainsi prêt, n'était certes pas conçu en réseaux neuronaux artificiels, mais il était en situation d'effectuer jusqu'à 200 millions d'opérations de calcul par seconde. C'était donc un duel entêté de la puissance de calcul engagée contre la haute intelligence humaine. Au moment où, à la fin de 1996, *deep blue* vainquit finalement Garri Kasparov,

⁷ J'utilise ici le terme « déchiffrage ou décryptage (*entschlüsselung*)», parce que je n'aime pas utiliser le terme « reconnaissance » (*Erkennung*) pour ce processus machinal.

grandit tout à coup l'intérêt porté à la question de savoir si un ordinateur pouvait encore « mieux penser » qu'un être humain.

« Intelligence artificielle »

IBM travaillait à son prochain super ordinateur qui était censé donner des réponses aux questions les plus simples. Il fut conçu en tant qu'une « intelligence artificielle », c'est-à-dire que les réseaux neuronaux artificiels devaient être élargis de manière telle qu'ils fussent en situation de se coltiner aux structures sémantiques.

Au contraire du simple réseau neuronal artificiel, qui ne « sait » rien des mots, lettres, couleurs et contextes de sens et dont la tâche dépend purement et simplement d'un seuil de potentiel adroitement affiné, on tentait ici d'édifier une « machine de recherche sémantique ». Une machine de recherche classique peut passer en revue à fond une banque de données à l'aide d'un mot dont elle recherche l'occurrence.

Par exemple, à la question :

- *Quand est-ce que mourut Olof Palmer ?*

Elle cherche à retrouver dans la banque de données des endroits où se présentent les mots « quand », « mourut », « Olof » et « Palmer ». Elle fournit donc, comme résultat, un grand nombre de sources d'informations qui ne sont pas importantes. Une machine de recherche sémantique « reconnaît » elle, qu'il s'agit du jour de la mort d'Olof Palmer et présente, dans le cas idéal que des sources d'informations qui sont pertinentes ici.

Elle est aussi en situation de reconnaître des variantes d'interrogations suivantes :

- *Quand Olof Palmer mourut-il ?*
- *Quel jour Olof Palmer fut-il assassiné ?*
- *Désigne la date de la mort de Olof Palmer.*

Et elle trouvera la même réponse à toutes ces questions. Idéalement elle peut aussi contourner avec cela une question dont la teneur est :

- *Désigne la date de la mort de Olof Palmer,*
- *Et fournir le type de mort et pas seulement la date.*

Un tel système a donc besoin d'un « savoir » sur des structures et des contextes sémantiques. Cela fonctionne d'une manière analogue au thésaurus du programme ELIZA (voir ci-dessus), et de maîtriser par dessus le marché des règles grammaticales de base et avec cela il peut en partie même « inférer » des « relations de sens » ou bien les dériver du « savoir » existant.

En 2011, *Watson* — ainsi s'appelait alors le programme d'IBM — fut finalement en situation de triompher dans l'émission américaine télévisée de *quiz* « *Jeopardy* ». Il avait une mainmise sur 100 Giga-octet en mémoire de textes, parmi lesquels des dictionnaires, encyclopédies et l'ensemble de *Wikipedia*. Jusqu'à aujourd'hui, *Watson* n'est capable de répondre certes qu'à des questions simples en anglais. Mais il est de plus en plus mis en œuvre pour des analyses de données et pour la gestion d'assistants numériques. [D'un autre côté *Watson* n'est encore pas *Holmes* non plus ! *ndt*]

Comme on le voit le domaine de recherche de « l'intelligence artificielle » s'articule aujourd'hui dans les domaines différents suivants :

- Senseurs, perception.
- Apprentissage automatique (Réseaux Neuronaux Artificiels RNA).
- Tirer des conclusions automatiques (logiques sémantiques).
- Travail du langage (*text-to-speech*, *speech-to-text*).

Au centre de l'imitation de l'intelligence, se trouvent cependant toujours comme auparavant les RNA qui sont aujourd'hui équipés de structures complexes et en partie jusqu'à plus de 100 couches intermédiaires entre l'entrée et la sortie. Avec cela des productions d'intelligence humaine se laissent imiter aujourd'hui dans une qualité surprenante. La reconnaissance de modèle par images, par exemple a tant progressé entre temps qu'à partir d'un pool de 1 million d'images, on peut reconnaître une personne recherchée. *Human parity* atteint ce score en 2015. Or *human parity* est défini de telle manière qu'un taux d'erreur de 5% est toléré, parce que c'est le taux d'erreur humaine maximum constaté dans ce domaine.

Un RNA doit être entraîné avec une grande quantité de phrases données. À l'occasion, il est important que les données soient les plus multiples possibles et qu'aussi des situations inhabituelles ou rares soient prises en compte. Cela pourrait irriter, par exemple les RNA, qui étaient entraînés dans la reconnaissance des panneaux de circulation pour l'automobile autonome qu'on dessine quelques points noirs sur divers

panneaux. Ou bien la reconnaissance de personnes peut être durablement irritée par la présence de fins pointillés sur l'image, qui sont pour nous invisibles, mais perturbent les machines durablement.

2016 : AlphaGo vainc le maître-Go mondial Lee Sedol

Que les ARN sont des programmes de reconnaissance de types de formes c'est ce que démontre la victoire remportée par *AlphaGo* sur Lee Sedol. Lorsqu'on modifia les couleurs du damier, de sorte qu'elles ne correspondaient plus au damier d'entraînement, le RNA perdit soudainement ses capacités. C'est pourquoi c'est aujourd'hui une question importante de la recherche de savoir comment pouvoir apporter aux RNA des productions de transfert humaines. Une amorce actuelle consiste à entraîner sur notre monde les bases du savoir des RNA, par exemple, aux structures sémantiques ou à ce que sont des arêtes et des angles. Un tel « savoir » est ensuite emmagasiné en mémoire dans certaines couches des RNA et se trouve ainsi à disposition, lorsqu'on entraîne les RNA à leurs tâches véritables. Un tel RNA conserverait alors ses capacités même en cas de changement de couleur du damier.

Forte ou faible intelligence artificielle

Étant donné que l'imitation d'intelligence se trouvant aujourd'hui à disposition par les machines ne peut que ponctuellement surpasser les productions de l'intelligence humaine, et qu'elle est donc toujours et encore bien loin de passer de manière satisfaisante le test de Turing, on a introduit une distinction en « forte » et « faible » intelligence artificielle. L'ia postulée comme « forte » est censée être capable dans quelques décennies de s'optimiser de manière autonome, de percevoir des émotions chez l'être humain et de coopérer avec lui « à la hauteur des yeux ». Elle doit aussi accomplir des prestations artistiques. Avec elle de nombreux problèmes se laissent résoudre ou pour le moins minimiser : dans la circulation automobile (conduite autonome), la recherche, dans le travail de la police (police prédictive), dans la médecine et ainsi de suite. Lorsqu'une telle ia est engagée dans le contexte de la construction et de l'administration urbaines, on parle alors de *smart city*. Une *smart city* peut gérer le flot de circulation de manière optimale, évaluer le comportement des citadins et reconnaître des déviations ou des infractions aux normes fixées, lorsque quelqu'un devient subitement malade ou criminel et y réagir de manière « adéquate ». Lorsque des robots sont gouvernés par l'aide de l'ia, ceux-ci sont capables aussi de hauts faits sportifs et en mesure de remporter la coupe du monde de football en 2050.⁸

Les prophéties de l'ia

Jusqu'à présent la « forte ia » n'est cependant restée encore qu'un postulat. Les capacités arrivent souvent bien plus tardivement que prévues et de plus sous leurs aspects partiels. L'informaticien Rodney Brooks se fait un plaisir de rompre l'enchantement de toutes ces « prophéties de l'ia ». Sur son site *web*⁹, on trouve beaucoup de matériau à ce sujet. Les méthodes qui sont utilisées par les prophéties de l'ia, sont classées par lui de la manière suivante :

1. L'ia est en même temps sur- et sous-estimée : les conséquences à court terme sont en général sur-estimées, celles à long terme sous-estimées.
2. L'ia est magique : ce que l'ia doit pouvoir tout réaliser à l'avenir apparaît si fantastique que d'un côté cela fascine et, de l'autre ce n'est pas contestable.
3. L'ia se voit imputer une compétence humaine : Lorsque l'ia accomplit une performance technique extrême, nous partons inconsciemment du fait qu'elle peut alors maîtriser toutes les productions quotidiennement importantes que nous, nous maîtrisons.
4. L'ia est présentée en « mots-valises » : « Penser » est, dans ce contexte, un tel mot-valise, avec lequel un contenu, tiré d'un contexte de sens humain, est détaché et replacé dans un contexte qui lui est complètement étranger.
5. Les prophéties de l'ia sont dominées par la potentialité de la foi : Des bonds soudains dans le développement technique se voient aussitôt extrapolés dans l'avenir.

⁸ Pour cela il existe chaque année en différent lieu une « *robotCup* » : une série d'équipes de recherche sont en lice dans diverses disciplines par le truchement de leurs robots qui ont à faire avec le football. Dans l'une se sont des robots « bipèdes » qui s'affrontent et on analyse alors leur capacité à se tenir solidement debout. Dans une autre, il s'agit de la coordination entre les joueurs et dans ce cas il ne s'agit pas de robots bipèdes. Les conditions de participation sont rendues plus difficiles chaque année afin que l'attrait subsiste d'améliorer continuellement les machines. À la fin on est censés ré-assembler ces diverses participations en une sorte d'équipe de footballeurs humanoïdes, qui est censée remporter la coupe du monde de ce sport en 2050. [Rappel : la notions de robot fut développée dans les années 50 et 60 dans les « romans » de l'écrivain de science fiction et biochimiste américain vulgarisateur, **Isaac Asimov** (voir sur Wikipedia) qui fut le premier à tenter d'imaginer les règles et principes de comportements entre robots et humains pour la sauvegarde de ces derniers (voir cet aspect dans l'article de Kai Ehlers qui suit). *Ndt*]

⁹ rodneybrooks.com

6. L'ia dans des scenarii hollywoodiens : dans les films de science-fiction, une ia est souvent complètement isolée et présentée dans notre temps présent sinon inchangé.
7. La vitesse de diffusion de l'ia est sur-estimée outre mesure : car aucun entrepreneur ne jettera aux orties, du jour au lendemain ses outils-machines préservés et ne les remplacera par une technique inexpérimentée dirigée par l'ia.

Lorsqu'à un moment quelconque, on tombe sur un de ces modèles d'argumentation, que ce soit chez les défenseurs et les critiques de l'ia, il faut être prudent.

La singularité — Évolution d'une mésintelligence de soi

Il peut foncièrement arriver que quelques années après on n'entende plus parler principalement de ses choses, parce que les buts promis — une imitation machinale de l'intelligence humaine, n'a pu être atteinte de ce côté. Ce ne serait alors jamais que le « troisième hiver de l'ia ».

Mais même dans ce cas, on peut partir du fait que les espoirs et les visions qui se rattachent aux recherches sur les RNA continuent de vivre jusqu'à ce qu'un jour, de nouvelles technologies se trouveront à disposition qui donneront à nouveau une impulsion à cette orientation de recherche.

Il faut espérer qu'il est devenu évident que les idées et visions techniques qui se trouvent derrière l'ia, sont déjà nées dans les années 1950, bien avant que l'on eût principalement une possibilité d'avancer dans cette direction. Elles disparaîtront seulement d'autant moins pour la raison qu'à brève échéance ces possibilités techniques sont inexistantes.

Comme objectif un pur paradis sur Terre

Je plaide fortement pour cette raison pour tenir séparées les réelles évolutions techniques des motivations profondes qui les sous-tendent. Celles-ci agissent comme donneuses d'impulsion dans l'humanité et poussent l'être humain dans une direction déterminée. D'une part, une imitation technique de l'être humain est censée être créée, une idée qui ne cesse de vivre déjà depuis de nombreux siècles. L'être humain deviendrait ainsi un dieu sur Terre. D'autre part, un monde technique devrait être créé qui, en ultime instance, repousserait totalement notre monde naturel et nous ferait cadeau d'une santé éternelle, d'une vie éternelle, débarrassée de toute contention. Et donc un paradis sur Terre dans lequel nous vivrions divinement.

Le progrès technique qui s'accélère sans cesse

Cette motivation qui cible la création d'un paradis terrestre avec l'aide technique, se développe et se différencie sans cesse. Elle émergea dans les années 1950 et certes lors d'un entretien entre John von Neumann et Stanislav Ulam. Ce dernier en a rapporté la teneur :

*« La discussion s'infléchit au sujet de l'accélération constante du progrès technique et des changements dans la manière de vivre qui crée l'apparence qu'elle aboutira un jour sur une **singularité** décisive dans l'histoire de l'humanité, après laquelle les conditions de vie que nous connaissons ne pourraient pas se poursuivre. »¹⁰*

Ici, pour autant que je sache, émerge pour la première fois le motif de l'accélération de l'évolution technique (la foi dans son caractère exponentiel). Tombe dans cette catégorie la « loi de Moore » de l'année 1965, d'après laquelle le nombre des composants des circuits électroniques intégrés doublerait tous les 18 mois. À laquelle s'oppose ensuite toujours la relative constance de la faculté de prestation du cerveau humain [à ce qu'il semble à première vue, *ndt*]. Il doit donc résulter un jour de cela qu'à un moment quelconque dans l'avenir, la technique « surpassera » l'être humain. Or ce moment est désigné comme la « singularité ».

Machines ultra-intelligentes

Une autre évolution de cette figure d'argumentation nous pouvons l'enregistrer chez I. J. Good. Il formula en 1965, la première fois que quelque chose comme une ia devait s'y glisser :

« Une machine ultra-intelligente serait définie comme une machine qui pût dépasser largement les facultés intellectuelles de tout être humain, fût-il pourtant intelligent. Étant donné que la construction justement de

¹⁰ Stanislas Ulam : *Tribute to John Neumann [Hommage à John Neumann]* : dans *Bulletin of the American Mathematical Society* 64/3, partie II, mai 1958, p.5 — (soulignement en gras de S.P.)
https://projecteuclid.org/download/pdf_1/euclid.ams/1183522369

telles machines est une de ces facultés intellectuelles, une machine ultra-intelligente peut encore construire de meilleures machines ; on en arriverait sans doute ensuite à une évolution explosive de l'intelligence, et l'intelligence proprement humaine en serait dépassée. La première machine ultra-intelligente est donc la dernière découverte que l'être humain a à réaliser. »¹¹

Une telle suggestion œuvre avant tout à la manière du caractère magique de l'imitation de l'intelligence humaine. Non seulement on postule qu'une machine serait pour cela en situation d'imiter l'intelligence humaine, mais encore qu'elle la surpasserait. D'une façon semblable, Vernon Vinge en 1993, dans son ouvrage *Technological Singularity [Singularité technologique]*, « pronostique » que dans un laps de temps de 30 ans [et donc au plus tard en 2023, S.P.] nous disposerions des moyens technologiques pour créer une intelligence supra-humaine. Peu après, l'ère de l'humanité s'achèverait. »¹²

Le même son de cloche résonne chez le Pr. Dr. Jürgen Schmidhuber (Directeur scientifique de l'*Instituto Dalle Molle Sull'Intelligenza Artificiale — IDSIA* — Lugano, CH) et président de la *Neural Network Solutions for Superhuman Perception and Intelligent Automation NNAISENSE*) qui dans le monde de la recherche sur l'ia prend une petite place mais garde une position-clef. Il couvre du regard le développement de la technologie de l'ia et il a suivi intimement le développement des micro-circuits électroniques intégrés :

« Lorsque je commençai, dans les années 1980, à travailler les réseaux neuronaux (ARN), les ordinateurs étaient un million de fois plus lents qu'aujourd'hui et nous ne pouvions mener alors que tout petites expérimentations de jeu avec nos premières machines neuronales fonctionnellement habiles « Deep Learning [« apprentissage profond » en anglo-saxon(tordu, car une fois au fond, il faut bien « remonter ») ndt] »

Mais il succombe aussi à la suggestion du développement exponentiel et conclut en conséquence :
« L'évolution d'une intelligence artificielle vraie est l'ultime chose importante que l'être humain peut encore produire. »¹³

La fin d'une culture purement humaine

Ici résonne pour la première fois un motif eschatologique, tandis que se chuchotent des messes basses au sujet d'une imminente fin de la culture humaine. C'est ce motif qu'a repris ensuite Raymond Kurzweil dans son ouvrage *The Singularity is near [la singularité est proche]* en 1998 [En allemand : « *Menscheit 2.0* »]. Selon Kurzweil on sera bientôt au moment où en 2045, on pourra pour ainsi dire « décharger » son cerveau, notoirement dans l'ordinateur [bien entendu pour ceux qui en auront encore gardé un, je suppose, ndt] Et avec cela les êtres humains seront éternels — pour le moins au niveau numérique. Le savoir et l'intelligence seront en mémoire à jamais, bien au-delà de la mort de chaque personne.

Dans son article *The law of Accelerating Returns [La loi des renvois accélérant]*, il présenta en 2001 la thèse que la loi de Moore n'est que le cas particulier d'une loi générale selon laquelle le changement technologique s'accélérait lui-même de manière exponentielle. Alors que dans les milieux informatiques « normaux », les discussions tourneraient autour de la loi de Moore qui perd lentement de sa validité, parce que dans l'intervalle la miniaturisation a atteint le niveau du domaine atomique, Kurzweil affirme pratiquement exactement le contraire. Dans sa compréhension de l'évolution de l'humanité-cosmique, il y a...

« ... un changement technique qui est si rapide et globalisant qu'il représente une rupture dans la structure de l'histoire de l'humanité. »¹⁴

Humanité plus ai ou seulement ai ?

Depuis les apologistes de la singularité ne discutent plus au sujet de savoir si et quand ceci aura lieu. Les spéculations tournent avant tout au sujet de savoir s'il s'agit de la naissance d'une « super-intelligence artificielle » qui mènerait à la suppression de l'humanité ou de son évolution ultérieure, si l'évolution d'après est un parcours purement technique, sous la forme de super-intelligences s'optimisant techniquement elles-mêmes et se continuant comme telles ou bien s'il s'agit d'un sorte de co-évolution lors de laquelle l'être

¹¹ Cité d'après *Wikipedia*, https://de.wikipedia.org/wiki/Technologische_Singularität [Dans un cas pareil: surtout, ne pas oublier de débrancher la prise, en partant ! ndt]

¹² Vernon Vinge : *Technological Singularity*, Department of Mathematical Sciences San Diego State University, 1993 — <http://mindstalk.net/vinge/vinge-sing.html>

¹³ Pr. Dr. Jürgen Schmidhuber : *Comment l'ia modifiera notre monde*, 17.1.2016 —

<https://www.xing.com/news/klartexte/wie-kunstliche-intelligenz-ki-unsere-welt-verandern-wird-386>

¹⁴ Ray Kurzweil : *The law of Accelerating Returns Essay* 7 mars 2001 — <https://www.kurzweilai.net/the-law-of-accelerating-returns>

humain continuerait aussi son perfectionnement des facultés par une sorte de fusion avec la super-intelligence technique (post-humanisme *versus* transhumanisme).

Le post-/transhumanisme comme religion nouvelle

La culmination précurseuse de ce développement complètement spéculatif et au plus vrai sens du terme méta-physique, c'est la fondation de l'église *The Way of the Future Church*. Ici, cette fleur de style du penser humain en arrive exactement au point où elle est, en vérité, à savoir, à une religion !

« Les choses en lesquelles nous croyons : »

« Nous croyons que l'intelligence n'est pas enracinée dans la biologie. Alors que celle-ci a fait évoluer un variant d'intelligence, il n'y a rien de spécifiquement inhérent à la biologie qui cause l'intelligence ; Éventuellement nous serions capables de la recréer sans utiliser la biologie et ses limitations. De là nous pourrions l'augmenter jusqu'au delà de nos limites biologiques (fréquence de calcul, vitesse et précision des copies de données, etc.) [...] **Nous croyons que la création d'une super-intelligence est inévitable** (principalement du fait qu'après l'avoir re-créée nous serons capable de la régler, de la fabriquer et de la mettre à l'échelle). Nous ne pensons pas qu'il puisse y avoir un moyen d'empêcher que cela arrive (ni que nous devrions même le vouloir) et que ce sentiment que nous devrions la stopper repose sur l'anthropomorphisme du 21^{ème} siècle (de manière semblable à la représentation d'un passé « qui n'est pas si éloigné » que le Soleil tourne autour de la Terre). »

Est-ce là le point final du matérialisme, une techno-religion ? Si cette pensée devait se répandre, on pourrait se voir carrément forcés en tant qu'êtres humains spirituels de défendre le noyau authentique du matérialisme contre ce nouveau genre de superstition nouvelle.

À suivre dans le prochain numéro :

Prophétie d'utilité personnelle | IA sottise | Domaines d'engagement de l'intelligence machinale aujourd'hui et dans le futur proche | Art : calculabilité du beau | Reconnaissance d'émotion | Autres champs de problèmes | Processus d'adaptation | Intelligence artificielle dans les vies économique, juridique et spirituelle | IA : La *Dreigliederung* sociale au banc d'épreuve | Multiplicité, participation, paysage | *Dreigliederung* sociale : changement de l'être humain, de la technique et de l'économie.

Sozialimpulse 4/2019.

(Traduction Daniel Kmiecik)

Stefan Padberg, est né en 1959 et a grandi à Fribourg-en-Brigau et a pris part aux mouvements de cet endroit des années 1970 et du début des années 1980. Il étudia la technique de l'information à Hambourg et travailla comme ingénieur. Depuis 2012, il exerce une profession indépendante et est actif comme programmeur du *Web*. La crise financière l'amena, en 2007, à s'occuper plus intensément des thèmes d'économie sociale et de l'idée de la *Dreigliederung*. Dans cette évolution, il collabora au parcours d'études d'évolution sociale à l'*Institut pour les questions sociales du présent* de Stuttgart et à l'édification de l'*Initiative Netzwerk Dreigliederung*. Stefan Padberg est par ailleurs directeur du cercle de travail Europe auprès de *Mehr Demokratie e.V.*

Courriel : post@futur3.org

Notes du traducteur :

- (a) (Pas d'équivalent ni en latin ni en français) mais je comprends ce terme comme « l'ensemble des fascinations drainées par l'ia ». Il restera donc en allemand et souligné en caractère italique.
- (b) Technique de « recherche incontrôlée » d'idées, utilisée par la publicité surtout par la publicité, en deux temps : un temps d'imagination libre et incontrôlée et un temps de critique acerbe. Le CNRS dans les années 80 du siècle passé a essayé de la développer dans ses laboratoires en France, mais la première phase de la technique, qui supprime totalement la hiérarchie administrative d'un laboratoire français ordinaire n'a jamais été tolérée par les directeurs de laboratoire. L'idée est donc disparue en même temps que la gauche en France, dans les années 90.
- (c) C'était donc clairement un concept « creux », sans substance spirituelle!