



## **Hubert L. Dreyfus, Martin Heidegger et les autres**

---

**Hubert L. Dreyfus fut le premier à critiquer l'IA cognitive et computationnaliste en convoquant la phénoménologie, de Husserl à Heidegger.**

**15 janvier 2022**

Catégorie : **IA & Neurosciences**

Tags : **algorithme, cognitivisme, corps, esprit, futur, philosophie, société**

Personnages : **Hubert L. Dreyfus, Emmanuel Faye, Edward Feigenbaum, Richard Greenblatt, Julian Feldman, Martin Heidegger, Alan Newell, Seymour Papert, Herbert Simon, Terry Winograd**

## L'IA à l'abordage

L'Intelligence Artificielle (IA) est sans doute la technique du siècle et son ambition sans limites : assister voire remplacer l'humain industriel, surveiller et faire la guerre, s'emparer du pouvoir cognitif... Le siècle s'ouvre ainsi sur la possibilité d'une véritable rupture de civilisation. Mais peu l'aperçoivent car l'IA avance triplement masquée : reléguée derrière notre double souci écologique et sécuritaire, drapée dans son opacité technique et philosophique, dérobée à notre enquête par les puissances numériques, politiques et militaires. Nous voici donc exposés à une technique radicale qui échappe à toute critique sérieuse, non pas tant au sujet de ses applications techniques, vaguement tempérées par des promesses éthiques et juridiques, qu'au sujet des *conditions-même de possibilité* d'une prétention toujours intacte : atteindre voire dépasser l'intelligence humaine.

L'IA a pourtant trébuché plusieurs fois depuis son acte de baptême par John McCarthy et ses collègues en 1956. Elle a traversé deux « hivers » (1974-1980 et 1987-1993) mais progressé inlassablement en tirant parti pour l'essentiel de la puissance des ordinateurs. Les algorithmes de 2020 ne sont pas beaucoup plus « malins » que ceux de 1960 mais l'abondance de données et la puissance de calcul sont sans commune mesure et drapent aujourd'hui ces algorithmes d'un semblant d'intelligence (le mouvement des poils dans les films animés par ordinateur a fait beaucoup de progrès, mais ce ne sont toujours pas de vrais poils...).

Ceux qui, comme moi, ont débuté leur carrière dans l'IA entre ses deux hivers, à la grande époque des « systèmes experts » et des premiers réseaux neuronaux artificiels opérationnels, avaient de l'ambition et croyaient sincèrement à la possibilité d'une authentique intelligence artificielle. Une pointe de phénoménologie (et même de bon sens) aurait pu tempérer notre élan pour ce qui n'était, au fond, qu'une boîte à outils

algorithmique. Les systèmes « intelligents » se multipliaient donc comme des petits pains mais au prix d’hypertélies et de rafistolages conceptuels et techniques toujours plus saugrenus. Ce n’était d’ailleurs pas tant le fait des ingénieurs qui réalisaient ces systèmes que des chercheurs eux-mêmes qui empilaient les théories sans jamais devoir ni même pouvoir rendre compte du bien-fondé de leurs idées (pour ne donner qu’un seul exemple qui parlera sans doute à certains, un summum de passage en force conceptuel est la fameuse « *grammaire générative et transformationnelle* » de Noam Chomsky). L’Amérique donnait des leçons en matière de cognition qui avaient la finesse d’un abordage de corsaires. Bref, le « cognitivisme » faisait des ravages et nous ne devons qu’à la loi de Moore et, encore aujourd’hui, le camouflage de ses ruses : les vessies algorithmiques passent pour des lanternes intelligentes à la vitesse de millions de milliards d’opérations par seconde.

C’est à peu près la raison pour laquelle les plus « sensibles » d’entre nous ont lâché l’IA et ses belles promesses, au sortir de son deuxième hiver, paradoxalement juste avant le premier « exploit » de la discipline : la fameuse victoire en 1997 du superordinateur Deep Blue d’IBM sur le champion d’échecs Garry Kasparov. Mais la mise en scène de cet événement avait en quelque sorte conforté notre décision (je souligne)<sup>1</sup> :

*La victoire de Deep Blue était considérée comme un symbole important, un signe que l’intelligence artificielle rattrapait l’intelligence humaine et pouvait vaincre l’un des grands champions intellectuels de l’humanité.*

Il suffisait de connaître les ressorts de Deep Blue pour comprendre qu’il n’y avait aucun frémissement au sujet de l’intelligence et qu’une fois

---

<sup>1</sup> Wikipedia – [Deep Blue versus Garry Kasparov](#)

encore se répandait cette sémantique de flibustier qui armait l'essor de l'IA. On m'a demandé récemment comment nous, praticiens, envisagions l'avenir de l'IA dans les années 1990. L'IA avait déjà survécu à toutes ses outrances en continuant à s'appeler, comme John McCarthy l'avait voulu, « *intelligence artificielle* ». Elle se frayait donc inévitablement un chemin vers le troisième millénaire sous cette bannière. Je percevais donc l'IA comme la technique d'un siècle à venir dominé par la *puissance*, mais aussi comme une technique *crépusculaire*. Mon avis n'a pas changé depuis et j'y reviens en conclusion.

### **L'intrusion de Hubert Lederer Dreyfus**

Un philosophe n'a jamais cessé de critiquer cette prétention de la discipline à *encoder* l'intelligence humaine et animale et surtout à confondre, sans beaucoup de précautions épistémologiques, encodage et « décodage » (violant ce que le mathématicien Reuben Hersh appelait la zéro<sup>ième</sup> loi de la modélisation : ne jamais confondre le phénomène et son modèle). Il s'agit de Hubert Lederer Dreyfus, un philosophe américain disparu en 2017. En 1965, alors qu'il était professeur au MIT, il publia la première charge sérieuse contre deux des principaux promoteurs de l'IA : Allen Newell et Herbert Simon. Dans un document au titre explicite, « *Alchemy and Artificial Intelligence* »<sup>2</sup>, Dreyfus contestait vigoureusement l'allégation principale derrière laquelle se rangeait et se range encore toute la discipline : l'intelligence doit progresser le long d'un « *continuum* » dont le point de départ est déjà déterminé comme étant l'*ordinateur* (j'examine anthropologiquement cette doxa continuiste dans

---

<sup>2</sup> Ce document passionnant est toujours disponible en ligne, pièce importante dans l'histoire de l'IA et plus largement des idées – Hubert L. Dreyfus – Décembre 1965 – [Alchemy and Artificial Intelligence](#)

[Une lecture de Philippe Descola](#)) ; atteindre l'intelligence humaine n'est donc qu'une question de temps et de moyens. Mais aucun argument sérieux ne vient soutenir cette pétition de principe. Autant prétendre pouvoir atteindre la lune en grimpant à l'arbre sous prétexte qu'il existe un continuum spatial franchissable entre les deux. Ainsi Dreyfus observe (je souligne, ajouts entre crochets)<sup>3</sup> :

*Au lieu d'essayer de faire usage des possibilités spécifiques de l'ordinateur, les chercheurs en intelligence artificielle – aveuglés par leurs succès précoces et hypnotisés par le postulat que la pensée est un continuum – ne se contenteront de rien de moins que la lune [ l'intelligence humaine artificielle ].*

À l'appui de ce constat, Hubert Dreyfus rapporte l'entrée en matière d'un ouvrage de 1963, « *Computer and Thoughts* », des deux chercheurs Edward Feigenbaum et Julian Feldman. Nous reprenons ici l'extrait cité tant il est représentatif de l'ambiance de l'époque et, plus largement, du système de croyances occidental :

*En ce qui concerne le continuum de l'intelligence suggéré par Armer, les programmes informatiques que nous avons élaborés n'en sont qu'à leurs tout débuts. Il est fondamental de poursuivre notre lancée en direction de ce jalon que constituent les facultés de l'intelligence humaine. Y a-t-il la moindre raison de penser que nous n'y arriverons jamais ? Absolument aucune. Pas la moindre preuve, pas le moindre argument logique, aucun théorème n'a jamais mis en évidence le moindre obstacle insurmontable le long de ce continuum.*

---

<sup>3</sup> Ibid.2 p.83

La pétition épistémologique des chercheurs ne peut évidemment pas être démentie, tout ce qui n'est pas démontré pouvant être réputé vrai<sup>4</sup>. Le credo « computationnaliste » de l'IA (l'isomorphisme de l'intelligence à un programme informatique) ne supportait aucune contradiction. La férocité de la charge du jeune Dreyfus répondait à celle du catéchisme ambiant. S'avancant à un parallèle entre l'IA et l'alchimie, qui reposait de la même façon sur un continuum de progrès menant au Grand Œuvre, il provoqua la vindicte des chercheurs. Herbert Simon l'accusa ainsi de faire de la « *politique* » et méprisa un travail qu'il qualifia sans nuance d'« *imbécillité* » (« *garbage* »). Dreyfus se souvient que ses collègues du MIT travaillant en IA essayaient de ne pas être vus en sa présence à l'heure des repas. Seymour Papert organisa un match d'échecs entre Dreyfus et un programme développé par Richard Greenblatt. Dreyfus perdit le match et Papert déclara : « *un enfant de dix ans peut battre la Machine mais la Machine peut battre Dreyfus* ». Herbert Simon lui conseilla de se calmer et de retrouver son sens de l'humour, etc. Tous ces sympathiques personnages étaient pourtant nos « héros » de l'époque. Un rayon intellectuel transatlantique et divin venait nous éblouir. Lire Dreyfus et exprimer des doutes n'allait pas de soi, même ici en France dans les années 1980. Après tout, ce n'était « que » de la philosophie, et la philosophie devait se recroqueviller face à la puissance probante de l'ordinateur (rien n'a vraiment changé mais aujourd'hui la responsabilité est clairement passée du côté des philosophes). Dreyfus raconte ainsi<sup>5</sup> :

---

<sup>4</sup> Ce n'est pas sans rappeler les modes « argumentatifs » utilisés aujourd'hui sur les réseaux sociaux, qui ne visent pas tant le sujet débattu – sans intérêt véritable – que la position propre du locuteur dans l'espace du débat : notoriété, défense d'un « territoire », exclusion des hérétiques...

<sup>5</sup> Hubert L. Dreyfus / Elsevier, Artificial Intelligence Volume 171, Issue 18 – Décembre 2007 – [Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian](#)

*Quand j'enseignais au MIT au début des années soixante, les étudiants du laboratoire d'intelligence artificielle venaient à mon cours sur Heidegger et disaient : « vous les philosophes, vous êtes restés dans votre fauteuil depuis 2000 ans et vous ne comprenez toujours pas comment l'esprit fonctionne. Nous, au laboratoire d'IA, nous avons pris le relais et nous réussissons là où vous avez échoué. Nous programmons des ordinateurs pour qu'ils fassent preuve d'intelligence humaine : résoudre des problèmes, comprendre le langage naturel, percevoir et apprendre ».*

Ces réactions outrées et condescendantes, parfois violentes, pouvaient s'expliquer en partie, au-delà du préjudice d'image à l'égard de seigneurs contestés en leurs royaumes par Dreyfus, par la menace que fait planer tout mécréant sur les financements des recherches en IA et notamment, dans les années 1960, sur ceux de la DARPA, l'agence américaine chargée de la recherche en matière de nouvelles technologies à usage militaire. Hubert Dreyfus aurait pu se contenter d'attendre tranquillement l'inévitable échec des ambitions computationnalistes (cet échec surviendra d'ailleurs rapidement et les gouvernements américains et britanniques couperont les crédits de recherche en 1974, déclenchant le premier hiver de l'IA). Mais Dreyfus, au-delà de ses attaques, ne pouvait pas rester silencieux car, ne se contentant pas de critiquer, il proposait déjà une voie à explorer, une voie qu'il développera toute sa vie et qu'il esquissait déjà dans son pamphlet de 1965.

### **Sortir du cadre**

En 1982, le Japon et son initiative FGCS (« *Fifth Generation Computer Systems* ») ont largement contribué au renouveau de l'intérêt pour la programmation logique tant décriée par Dreyfus. Mais cette renaissance portait plus sobrement sur la « connaissance » (« knowledge

») et le « traitement de la connaissance » au moyen de la programmation logique plutôt que sur l'intelligence en général. L'IA avait apparemment baissé d'un ton. Il n'était plus question de poursuivre la progression sur ce continuum menant à l'intelligence humaine mais plus modestement de doter l'ordinateur de capacités de *raisonnement logique* dans des domaines très précis : diagnostic médical, pilotage de hauts-fourneaux, liquidation de retraites, surveillance des circuits secondaires de centrales nucléaires, etc. Ces « systèmes experts » ont connu un certain succès et ont permis à l'IA de trouver un premier élan industriel, l'installant ainsi progressivement dans le paysage économique.

Mais, en même temps, la recherche en intelligence artificielle pure ou « générale » ne lâchait rien et poursuivait inlassablement sa progression comme un discret feu de tourbe. A l'époque, tout ingénieur en IA était nécessairement contaminé par ses émanations, ne serait-ce que parce que la plupart des sociétés privées étaient codirigées par des chercheurs liés aux laboratoires. Si le premier hiver avait calmé les plus fanatiques, le milieu de l'IA poursuivait sa divine mission : fabriquer l'équivalent, à terme, d'un humain artificiel. Avant d'être une science, l'IA est bien une idéologie<sup>6</sup> !

J'avais parcouru à l'époque les ouvrages de Dreyfus mais une simple lecture, même répétée, même convaincante, ne suffit pas à « sortir du cadre ». Ce fut la pratique quotidienne, pendant plusieurs années, de ces systèmes de raisonnement logique qui ont progressivement sapé ma foi dans ce continuum dénoncé par Dreyfus. Je commençais à mieux deviner les conditions de possibilité d'une intelligence authentique (pas nécessaire humaine) mais elles se dessinaient sur un horizon tellement lointain que

---

<sup>6</sup> Notons qu'un article publié dans l'excellent Wired est intitulé [AI is an Ideology, Not a Technology](#). Mais le terme est utilisé dans un sens différent. Il est question dans cet article de l'*utilisation* idéologique de l'IA et non pas, comme le titre et nous-mêmes le suggérons, de l'IA en tant qu'idéologie.

le temps des ingénieurs – le mien en tout cas – n’était à l’évidence pas venu : passer de l’alchimie à la chimie ne serait pas l’affaire d’une seule génération. Je n’ai finalement rien manqué : les progrès de l’IA algorithmique ne s’inspirent que très vaguement l’intelligence humaine, celle que le bon sens entend et que la philosophie désigne.

À force d’emboîter des théories fumeuses pour bricoler des systèmes dont l’intelligence se dérobaît sans cesse, il apparaissait que ce que j’attendais naïvement était une manifestation d’authentique *autonomie* de ces systèmes, une « naissance » en quelque sorte. Mais quelle *raison* aurait un système artificiel de m’*échapper* ? Est-il possible d’encoder ce *motif* ? Peut-il à défaut *émerger* d’un dispositif algorithmique ? Etc.

### **L’intuition dreyfusienne**

Aujourd’hui, les travaux de Dreyfus sont largement partagés et considérés comme des contributions philosophiques importantes pour l’IA. Si l’on devait résumer d’une formule l’intuition du philosophe à laquelle je finis par me ranger, c’est qu’*un organisme n’est « intelligent » qu’à condition d’avoir du « souci » à se faire* (à ma connaissance, Dreyfus n’utilise pas exactement ces termes). Or, nul programme informatique, aussi complexe soit-il, ne se fait évidemment le moindre souci. Ce propos quelque peu sommaire constitue le premier fil de la pelote dreyfusienne.

Poursuivant cette voie de l’intelligence soucieuse, donc *incarnée* et dans le monde, Dreyfus lit Maurice Merleau-Ponty mais surtout Martin Heidegger dont il est un exégète réputé aux États-Unis. En 1990, il publie ainsi un commentaire de « *Être et temps* » intitulé précisément « *Being-in-the-world* »<sup>7</sup>, traduction anglaise du terme heideggérien « *In-der-*

---

<sup>7</sup> Hubert L. Dreyfus / Bradford – 1990 – *A Commentary on Heidegger's Being in Time, Division I*

*Welt-sein* » (« *Être-dans-le-monde* »). Sans affronter le rude versant philosophique de ce terme, il annonce déjà la prise critique de Dreyfus : l'intelligence ne vient qu'à cet être-dans-le-monde qui connaît les tracas de l'*existence*. Cette posture philosophique légitime au sujet de l'IA a sa bannière intitulée – c'est presque incroyable – « *Heideggerian AI* » (HAI) ! Précisons toutefois que si Dreyfus était le premier passeur de la philosophie heideggérienne dans les milieux computationnalistes américains, nous devons à la philosophe américaine Beth Preston l'inspiration plus concrète de ce courant vers la fin des années 1980<sup>8</sup>.

### **Idéologies**

Chacun peut comprendre intuitivement cette voie dreyfusienne. L'homme n'est pas cet animal doué d'une raison qui guide chacun de ses pas. Le cerveau ne calcule pas sur une *représentation* du monde comme un navigateur sur une carte GPS, calcul et représentation que l'on pourrait continûment améliorer pour, munis d'ordinateurs obéissant à la loi de Moore, atteindre voire dépasser l'intelligence humaine. Il existe plutôt une forme de couplage réciproque et « intéressé » entre le monde et l'organisme et qui provoque l'émergence de ce que *nous appelons* l'« intelligence » (voir aussi [Francisco Varela l'hétérodoxe](#)). Mais quoique soient la pensée, l'intelligence, la conscience et tous ces phénomènes de l'esprit, la condition de représentation de leur « essence », y compris sous l'apparence de ce couplage convaincant, est l'existence d'un socle idéologique. En effet, l'intelligence n'est pas le caractère d'une planète ou d'un virus mais de nous-mêmes. Le seul moyen d'échapper à cette

---

<sup>8</sup> Voir à ce sujet le texte de Terry Winograd (traduit par Jean Lassègue), célèbre professeur d'informatique à l'université Stanford : Terry Winograd / *Intellectica*, 1993/2, 17, pp. 51-78 – Septembre 1989 – [Heidegger et la conception des systèmes informatiques](#)

réflexivité c'est une réification de la pensée que seul le langage, donc l'idéologie, permet.

L'IA serait-elle donc à ranger du côté des sciences humaines ? Elle y échappe encore car aujourd'hui, loin des postures quelque peu agressives des fondateurs et tout en persévérant dans sa doxa informationnelle et computationnelle, l'IA cherche systématiquement ses critères de vérité dans le domaine apparemment moins idéologisé des neurosciences. Hubert Dreyfus lui-même, ainsi que tous les chercheurs et penseurs du courant de l'intelligence incarnée (citons au moins Michael Wheeler et Erik Rietveld) suivent eux-mêmes ce principe, cherchant à valider leurs modèles de l'« être-dans-le-monde » dans le comportement d'assemblées de neurones. L'IA a ainsi transmis aux neurosciences le flambeau de la sémantique continuitiste des années 1960, sidérant *enfin* tous les Dreyfus, c'est-à-dire tous les philosophes. Ainsi Stanislas Dehaene, célèbre neuropsychologue aujourd'hui à la tête du conseil scientifique de l'Éducation Nationale, peut tranquillement déclarer que nos enfants sont des « *super-ordinateurs* » à qui il faut fournir les « *données* » dont ils ont besoin<sup>9</sup>, ou encore que (je souligne le vocabulaire continuitiste)<sup>10</sup> :

*Nous sommes tous dotés d'une machine cérébrale extraordinaire qui dépasse les ordinateurs, l'intelligence artificielle... Car le cerveau humain déploie des algorithmes, des théories scientifiques, qui n'ont pas été imitées jusqu'alors par les machines.*

L'obsession occidentale pour la pensée calculante trouve toujours un chemin...

---

<sup>9</sup> Isabelle Boucq / La Tribune – 22 octobre 2018 – ["Nos enfants sont des super-ordinateurs" Stanislas Dehaene, Collège de France](#)

<sup>10</sup> Delphine Bancaud / 20 minutes – 5 septembre 2018 – [Stanislas Dehaene: «Nous sommes dotés d'une machine cérébrale qui dépasse les ordinateurs»](#)

## Heidegger, non mais si !

Si Hubert Dreyfus a combattu toute sa vie l'idéologie computationnaliste et sa conception de l'intelligence comme relevant d'une combinatoire de symboles, sa « passion » pour l'interprétation heideggérienne vaut-elle idéologiquement mieux ? On ne peut pas ignorer, en France tout au moins, les vives controverses au sujet de la personne Heidegger et de ses liens historiques et surtout intellectuels avec le nazisme (d'ailleurs, à ma connaissance, l'expression « intelligence artificielle heideggérienne » n'existe pas de ce côté de l'atlantique). Des débats houleux ont été notamment alimentés par les interprétations fort documentées du philosophe Emmanuel Faye et furent ranimés par la publication, à partir de 2014, des fameux « *Cahiers noirs* ». Je reste pourtant convaincu que, pour affronter la dure question de la nature de l'intelligence et de ses éventuels mécanismes, il *faut* tôt ou tard affronter l'héritage philosophique de Martin Heidegger. Ceux qui en France se bouchent le nez perdent encore une occasion de *nous* faire participer aux grandes questions du siècle comme celle de l'IA et, à l'horizon, celle du transhumanisme. La prudence exige en même temps de faire face aux interprétations les plus sombres des concepts philosophiques disséminés par Heidegger et dont Emmanuel Faye prétend qu'ils sont pour le moins équivoques, voire qu'ils armaient un cheval de Troie de l'idéologie nazie. Quoiqu'il en soit, à l'heure où l'on se préoccupe d'éthique, l'IA ne peut guère ignorer ce qu'elle met véritablement en jeu ni donc contourner Heidegger et sa « syntaxe » conceptuelle éventuellement décontaminée de sa « sémantique » historique.

Les chercheurs eux-mêmes n'explicitent guère, ni peut-être ne perçoivent, les bases idéologiques de leurs travaux. Au moins, à l'époque des premières charges menées par et contre Dreyfus, la bataille faisait du bruit et dégagait des flammes. On s'écharpait à visage découvert.

Désormais, les « croyances » sont directement implémentées par des somnambules dans des algorithmes opaques. C'est ainsi que l'IA s'enfonce désormais sans effort dans les angles morts d'une civilisation crépusculaire, la « *Zivilisation* » technocratique et matérialiste décrite par Oswald Spengler au début du siècle dernier. Un signe contemporain de ce crépuscule ? Un seul : l'ordinateur a déjà remporté la partie. Hubert Lederer Dreyfus et nous tous sommes déjà échec et mat dans quelques coups.