



HAL
open science

**L'IA, un artefact technologique porteur de promesses
d'amélioration et riche de ses zones d'ombres,
avant-propos éditorial au dossier "L'IA : raison &
magie", Quaderni, 2022/1, n°105.**

Thierry Ménissier

► **To cite this version:**

Thierry Ménissier. L'IA, un artefact technologique porteur de promesses d'amélioration et riche de ses zones d'ombres, avant-propos éditorial au dossier "L'IA : raison & magie", Quaderni, 2022/1, n°105.. Quaderni, 2022, 105, pp.9-18. 10.4000/quaderni.2208 . halshs-03583065

HAL Id: halshs-03583065

<https://shs.hal.science/halshs-03583065>

Submitted on 21 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thierry Ménissier
Thierry.menissier@univ-grenoble-alpes.fr

Preprint de l'avant-propos éditorial du dossier d'articles
« *L'IA, entre raison et magie* »
paru in *Quaderni*, n°105, Hiver 2021-2022
<https://www.cairn.info/revue-quaderni-2022-1.htm>

Résumé du dossier

L'intelligence artificielle renvoie à la manière dont les machines algorithmiques ordonnent et régulent tout ce qui est susceptible d'être capté et traduit en données. En cours de déploiement massif, elle désigne un système sociotechnique global dont les applications sont innombrables. Mais l'efficacité de ce système ne vaut pas explication. L'observation des faits révèle des zones d'ombre sur la nature de l'interaction humain-machine. Les promesses d'amélioration (de l'efficacité, de l'existence et même de la vie humaine) par l'IA apparaissent corrélatives d'un considérable impensé social, ce qui ouvre un espace pour des hypothèses originales et l'invention de nouveaux cadres interprétatifs pour comprendre le phénomène IA.

Avant-propos

« *L'IA, un artefact technologique porteur de promesses d'amélioration et riche de ses zones d'ombres* »

Ce dossier réunit quatre contributions consacrées aux relations entre le système technique génériquement appelé « intelligence artificielle » et les modes de vie individuels ou collectifs contemporains qui constituent le cadre social de son déploiement. Dans ces contributions, l'IA est simultanément envisagée sous deux angles de vue qui peuvent de prime abord sembler contradictoire mais qui se trouvent l'un et l'autre appelés par la situation historique actuelle : d'une part, elle est considérée comme une des pointes avancées de la technologie contemporaine, peut-être susceptible, du fait de sa puissance et de ses champs d'application variés, de transformer profondément les manières de vivre et de penser héritées du passé (pour donner un seul exemple, l'hybridation humain-machine inclut aujourd'hui des algorithmes et conduit peut-être au transhumanisme). Mais d'autre part, comme les auteurs adoptent un certain recul historique, sa réception est envisagée comme un cas parmi d'autres d'un dialogue entre l'humain et les artefacts utiles qu'il engendre, favorisant un type d'interrogation dont les origines remontent probablement au début de l'histoire de techniques. L'IA, vectrice de situations inouïes et même de cas-limites, stimulerait donc un questionnement, qui, à défaut d'être banal, peut sembler classique. Les contributions réunies

ici permettent d'élucider ce début de paradoxe. On notera que les textes rassemblés relèvent d'une certaine variété disciplinaire, regroupant la philosophie politique, l'histoire et la philosophie des techniques, les sciences de l'information et de la communication et l'anthropologie. Ils ont été produits dans le contexte de la recherche menée au sein de la chaire « éthique & IA » de l'Université Grenoble-Alpes¹. Cette variété disciplinaire, avant même de formuler l'ambition de se positionner dans une perspective identifiée comme « éthique » à propos de l'IA, est nécessitée par le type de questionnement aujourd'hui proposé par l'essor de ce système technique particulier.

En effet, c'est en premier lieu la signification même de l'expression « intelligence » rapporté à l'IA qui interroge. Elle renvoie aux capacités des machines qui coordonnent de manière globale le calcul algorithmique, ces véritables dispositifs systémiques de production de données, qui calculent à partir des faits enregistrés par des capteurs, trient ces faits et les ordonnent, mais aussi apprennent d'eux et finalement en retour les « régularisent », et peuvent espérer les contrôler. Fruit de l'emploi systématisé de l'informatique, l'IA en reprend les concepts fondamentaux. Ainsi qu'on l'a remarqué, pour que l'informatique soit possible, les notions d'algorithme, de machine, d'information et de langage, différentes les unes des autres par leur histoire et même leur signification, ont été associées de manière originale². Pour autant, aujourd'hui l'IA demeure probablement davantage que de « l'informatique avancée » qu'une véritable intelligence. Et si, bien entendu, il serait périlleux d'entreprendre de définir l'IA d'une manière qui serait univoque, compte tenu de la grande extension de la notion, il apparaît possible de la comprendre en regard de sa genèse et de son évolution.

Du point de vue de sa genèse, d'abord : l'expression « *artificial intelligence* », la chose est bien connue, est apparue en 1956 à la conférence de Dartmouth au New Hampshire (États-Unis), organisée par J. McCarthy, C.E. Shannon, Marvin L. Minsky et N. Rochester. Si l'idée initiale était d'examiner comment des tâches différentes pourraient être accomplies par des programmes informatiques, les organisateurs de la conférence de Dartmouth ont également abordé diverses questions autour de l'idée de *machine pensante* : la première question « comment simuler la pensée et le langage grâce à des règles formelles ? » a engagé cette deuxième : « Comment doter une machine de capacité d'apprentissage automatique ? », qui a elle-même conduit à élaborer l'idée de machines de plus en plus autonomes dans les fonctions

¹ La chaire de recherche « éthique & IA », abritée par l'Institut de Philosophie de Grenoble, est soutenue par l'institut pluridisciplinaire en intelligence artificielle MIAI@Grenoble Alpes (ANR-19-P3IA-0003). <https://www.ethics-ai.fr/>

² Gilles Dowek, « Les origines de l'informatique », *Cahiers philosophiques*, 2015/2, n° 141, p.7-15.

qu'elles sont vouées à effectuer. Bien que la conférence de Darmouth n'a pas accouché de résultats de recherche immédiats, elle a lancé une impulsion dont l'effervescence contemporaine sur l'IA est l'héritière lointaine, et a permis de délimiter les contours d'une communauté de recherche autour des problématiques évoquées. Toutefois, le développement de l'IA a été inclus dans un projet de société cohérent car en phase avec la perspective moderniste de la rationalité et de la science (esquissée par Galilée et Descartes, validée par Newton et Leibniz). Ce projet de société, Norbert Wiener (1894-1964), l'avait préparé dans les années 1940 en imaginant la cybernétique comme science du fonctionnement de l'esprit humain³.

Du point de vue de son évolution, ensuite : deux grandes tendances techniques, qui se succèdent chronologiquement, ont orienté le développement de l'IA. La première s'apparente à un traitement de l'information, c'est à dire un ensemble de symboles obéissant à des règles. Dans ce cas, la machine cherche à modéliser ces symboles et à reproduire, dans le code, les règles qui les gouvernent. Cette vision a connu son heure de gloire dans les années 1980 et 1990, à travers les « systèmes experts ». La seconde, connue depuis de nombreuses années mais qui s'est développée de manière spectaculaire au cours de la dernière décennie, relève de l'apprentissage automatique (*machine learning*). Elle s'appuie sur des modèles mathématiques et des statistiques pour donner aux ordinateurs la capacité d'apprendre à partir d'exemples. Avec cette approche, l'idée n'est donc pas de *programmer* une machine, mais de *l'entraîner*. Par analogie avec le cerveau humain, ces systèmes utilisent des « réseaux de neurones » artificiels. Yann LeCun, Yoshua Bengio et Geoffrey Hinton ont dans les années 2000 développé une nouvelle technique d'apprentissage automatique, qu'ils ont appelée *deep learning*, ou (« apprentissage profond »), elle-même rendue possible par l'augmentation de la puissance des algorithmes et par l'afflux massif de données (*big data*)⁴.

Cette dernière évolution conduit, en second lieu, à considérer que l'« intermédiation algorithmique »⁵ d'aujourd'hui représente un moyen global de se rapporter à la réalité, certes, mais également un artifice brut qui gagne à être accompagné par la réflexivité propre aux sciences humaines et sociales. Un tel dessein impliquerait de déterminer avec précision quels sont les modèles dont on se sert pour qualifier l'intelligence dans l'expression « IA ». Ce point

³ Norbert Wiener, *Cybernetics or control and communication in the animal and the machine*, New York, Willey & Sons, 1948.

⁴ Yann LeCun, Yoshua Bengio et Geoffrey Hinton, « Deep learning », *Nature*, vol. 521, mai 2015. Ces trois auteurs ont reçu le Prix Turing 2019 pour leur apport à la science informatique.

⁵ Stéphane Grumbach, « Qu'est-ce que l'intermédiation algorithmique ? », *Bulletin de la société informatique de France*, numéro 7, novembre 2015, p. 93–111.

porte sur le statut épistémologique et ontologique de la calculabilité, de l'incomplétude et de l'incertitude, en relation avec les thèmes de la conscience, de la nature de la connaissance humaine toutes facultés confondues, ou encore du rapport entre rationalité, perception, mémoire et réflexivité⁶. Ce dessein concerne également l'explicitation des finalités poursuivies par le développement de l'IA, ainsi que par la détermination des valeurs qui sous-tendent ce dernier. La question est de savoir si le (ou les) paradigme(s) d'« intelligence » typiques de l'IA donnent ou non un *nouveau sens* au projet des sociétés modernes. Cette question peut se décliner de manière plus fine en ces termes : comment l'IA transforme-t-elle les pratiques, les usages et les représentations ? La relation qui s'installe de la sorte entre l'observation des transformations observables et la réflexion éthique et politique engendre un questionnement qui peut être formulé de la sorte : *quel type de société cherche-t-on à promouvoir avec l'IA ?* L'usage massif des algorithmes dans tous les secteurs de l'activité humaine renvoie en effet à un projet de société ancré dans une certaine définition de la modernité, un projet dans lequel la « raison statistique », ancêtre de l'algorithmique, joue un rôle majeur et supprime des formes plus classiques de conception de la décision publique, ainsi que l'ont établi plusieurs études de référence⁷. En termes d'éthique publique, le projet de la raison statistique a pris, avec l'essor de l'IA, un tour radical qui peut s'exprimer comme « gouvernementalité algorithmique »⁸ et en regard de ce qu'on a récemment nommé la « menace de l'algocratie »⁹. Or, de telles évolutions sont elles-mêmes susceptibles d'engendrer des effets tels que l'émergence de formes de contrôle automatique des comportements ou, sur le plan des valeurs, la primauté de l'ordre et de la sécurité sur le respect des libertés publiques et de la *privacy*.

En travaillant ces questions, les contributions du dossier entendent saisir le projet de la raison algorithmique à travers ses expressions concrètes. Les inventions technologiques contemporaines sont alors conçues comme des dispositifs sociotechniques compris au sein de sociétés d'innovation, où les offres de nouveaux produits et services technologiques proposées par les firmes innovantes sont validées par les choix et les pratiques d'utilisateurs individuels

⁶ Voir par exemple les études réunies dans le volume de synthèse coordonné par Jean-Marie Chevalier et Benoit Gaultier, *Connaître. Questions d'épistémologie contemporaine*, Paris, éditions d'Ithaque, 2014.

⁷ Alain Desrosières, *La Politique des grands nombres. Histoire de la raison statistique*, Paris, La Découverte, 1993 ; Alain Supiot, *La gouvernance par les nombres*, Paris, Fayard, 2015 ; Thomas Berns, *Gouverner sans gouverner. Une archéologie politique de la statistique*, Paris, P.U.F., 2009.

⁸ Antoinette Rouvroy et Thomas Berns, Thomas, « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ? », *Réseaux*, 2013/1 n° 177, p. 163-196.

⁹ Voir John Danaher, « The Threat of Algocracy: Reality, Resistance and Accommodation », *Philosophy & Technology*, 29, p. 245-268 (2016).

(consommateurs et citoyens) et collectifs (entreprises et institutions clientes)¹⁰. Ce sens n'a jamais été suffisamment qualifié, du moins à observer l'indécision récurrente en ces matières, où les auteurs se demandent, à propos des artefacts, robots et IA, si ces nouvelles entités sont plutôt des choses ou plutôt des dieux, en faisant souvent référence tant aux concepts philosophiques qu'à ceux de la tradition théologique¹¹. Autant d'éléments qui suggèrent qu'une approche approfondie en termes de philosophie de la technique, appuyée sur la connaissance sociale, historique et anthropologique, s'avère nécessaire afin d'espérer disposer d'une éthique appliquée l'IA¹².

Dans ce dossier, chaque auteur développe un angle de vue personnel, mais avec ce point commun de considérer que l'intelligence artificielle, étant conçue par des intelligences naturelles dans des contextes précis et en vue de certains usages, est fondamentalement une œuvre humaine porteuse de valeurs sociales. Ce qui signifie qu'aucun système d'IA ne saurait se prévaloir de la neutralité ou de l'objectivité qu'on veut parfois prêter aux objets techniques, et de ce point de vue il n'existe pas d'« autonomie » possible pour l'intelligence artificielle. Pour autant, les valeurs associées aux artefacts ne sont ni explicites ni claires. Elles apparaissent même caractérisées par une forme d'indétermination, en vif contraste avec les promesses dont elles sont porteuses selon les prescripteurs d'IA ou pour les utilisateurs. Et ce n'est pas parce que certains dispositifs touchent à l'intime ou à l'institution, l'un et l'autre domaines qui peuvent sembler indiscutables car ou bien intuitivement évident (pour le premier) ou bien durablement installé (pour le second), que les choses sont d'elles-mêmes plus nettes ou aisées à qualifier.

Ainsi, en braquant son attention sur le phénomène social de l'augmentation technologique des personnes handicapées, Alexandre Bretel, souligne le fait qu'il existe, au sein de l'expérience prothétique du corps vulnérabilisé, à la fois un processus d'hybridation consentie de l'humain avec la machine et de larges zones impensées, où se déploie un imaginaire flou soustrait à l'évaluation éthique. De manière comparable, dans le cadre d'une exploration de la thématique de l'aliénation technique rapportée à l'IA, et en reversant les préjugés cartésiens qui

¹⁰ Sur le système de l'innovation, voir Benjamin H. Bratton, *Le Stack. Plateformes, logiciel et souveraineté* (éd. originale 2016), trad. C. Degoutin, Grenoble, UGA Éditions; Thierry Ménissier, « Innovation et histoire. Une critique philosophique », *Quaderni*, n°91, Hiver 2016, p. 43-55.

¹¹ Voir Norbert Wiener, *God & Golem Inc. Sur quelques points de collision entre la cybernétique et la religion* (1964), trad. C. Romana et P. Farazzi, Paris, Éditions de l'Éclat, 2016 ; Philippe Breton, *À l'image de l'homme, du Golem aux créatures virtuelles*, Paris, Éditions du Seuil, 1995 ; Mark Alizart, *Informatique céleste*, Paris, PUF, 2017 ; Alexei Grinbaum, *Les Robots et le mal*, Paris, Desclée de Brouwer, 2018 ; Nicolas Spatola & Karolina Urbanska, « God-like robots: the semantic overlap between representation of divine and artificial entities », *AI & Society*, 35, 2020 p. 329-341.

¹² Peuvent d'ailleurs être mobilisées, afin de mener à bien un tel projet, les œuvres de penseurs aussi différents que Lewis Mumford, Gilbert Simondon, Bernard Stiegler et Andrew Feenberg.

poussent à considérer l'artefact comme un pur produit de la raison au profit de l'hypothèse que la machine, comme toute œuvre technique, relève de la vie, Tyler Reigeluth prend en considération le malaise engendré par le flou qui entoure la qualification de l'effet que nous fait les machines complexes inventées pour servir les humains. Si bien que « saisir la technique dans sa continuité avec l'activité vitale » conduit à envisager que la machine est porteuse de normes aussi invisibles que prégnantes, qu'il est nécessaire d'objectiver. Un tel renversement de perspective conduit à tester l'hypothèse que les machines algorithmiques sont caractérisées par une efficacité de type magique. Fabienne Martin-Juchat, en fonction d'une autre perspective, émet une hypothèse comparable : les objets communicants animés par l'IA, tels qu'ils sont aujourd'hui conçus, puis « mis en marché » et enfin adoptés par des usagers, sont sous-tendus par des valeurs originales pour lesquelles il faut concevoir le cadre d'un « néo-animisme ». En d'autres termes, la compréhension actuelle de l'usage des systèmes d'IA – usage pourtant en voie de devenir massif – demeure incomplète voire partiellement impossible si l'on ne l'inscrit pas dans un horizon de sens original. Si une telle interprétation s'avère audacieuse, c'est qu'elle affronte ce qui est contre-intuitif compte tenu des représentations traditionnelles à propos des machines, même pour celles considérées comme « intelligentes ». Elle est perturbante dans ses conséquences, car elle tend à modifier ce qu'on doit penser pour qualifier les relations entre les humains et les artefacts socio-techniques, d'une part, et entre la pensée rationnelle et les besoins de niveau spirituel, de l'autre. Enfin, Thierry Ménissier, observant d'abord la confusion qui s'installe entre la fiabilité de l'expertise algorithmique et la notion de confiance, s'interroge sur les conséquences de l'adoption massive des systèmes d'IA par les services publics. Les fonctions régaliennes des institutions, si elles ont depuis des temps très anciens été servies par des dispositifs sociotechniques nombreux et variés, courent aujourd'hui un risque nouveau, sous l'effet de l'apparition d'une forme d'autorité des machines.

En d'autres termes, pour tous les cas de déploiement de l'IA dans la société que les auteurs des contributions envisagent, l'observation minutieuse des faits conduit à la nécessité de recourir à des hypothèses originales et à celle d'inventer de nouveaux cadres interprétatifs, ce qui permet de révéler des significations inouïes et éventuellement dérangeantes, dans un effort qui s'avère absolument nécessaire si l'on veut repenser de manière efficace la partition actuelle entre l'humain et ses artefacts, au cœur de l'activité sociale où leurs relations se jouent.

Sommaire du dossier :

1. Alexandre Bretel : « *Corps reconfigurés ou corps augmentés ? Techno-imaginaire de la figure du cyborg chez les personnes en situation de handicap* »

2. Tyler Reigeluth : « *Le rapport magique à l'intelligence artificielle, ou comment vivre avec l'aliénation technique* »
3. Fabienne Martin-Juchat : « *Sur le néo-animisme technologique à l'ère de l'engouement pour l'intelligence artificielle* »
4. Thierry Ménissier : « *Jusqu'où l'institution peut-elle être augmentée ? Pour une éthique publique de l'IA* »