



HAL
open science

Intelligence artificielle, éthique & vieillissement

Thierry Ménissier

► **To cite this version:**

Thierry Ménissier. Intelligence artificielle, éthique & vieillissement. 4e Journée “ Silver Eco ” organisée par le Technopôle Alpes Santé à Domicile & Autonomie (TASDA) et le pôle de compétitivité Minalogic, Grenoble: “ La maison de demain: quelles briques innovantes? ”, Nov 2021, Grenoble, France. halshs-03460366

HAL Id: halshs-03460366

<https://shs.hal.science/halshs-03460366>

Submitted on 1 Dec 2021

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.

« Intelligence artificielle, éthique & vieillissement »

Thierry Ménissier

Professeur, Philosophie, chaire éthique&IA, MIAI Université Grenoble Alpes

Conférence d'ouverture de la 4^e Journée « Silver Eco » organisée par le Technopôle Alpes Santé à Domicile & Autonomie (TASDA) et le pôle de compétitivité Minalogic, Grenoble : « La maison de demain : quelles briques innovantes ? », Grenoble, 25 novembre 2021.

Résumé : Dans cette conférence d'ouverture de la journée d'étude consacrée à l'IA dans la "maison de demain" en regard des problématiques de vieillissement et de "silver économie", il s'agit de réfléchir à la nécessaire sagesse qui doit accompagner le développement des technologies d'IA et leur implémentation dans les services innovants. Dans cette tâche, la pensée d'Aristote apparaît inspirante, tant du point de vue de son apport sur la notion de technique qu'en regard de sa définition de l'autonomie humaine. Elle permet également de fonder la distinction entre "augmentation" et "stimulation" de l'humain par les technologies.

Mots-clés : Intelligence artificielle, innovation, éthique, vieillissement, silver économie

Bonjour, je suis heureux d'intervenir aujourd'hui sur la sollicitation du « centre expert des innovations pour le bien-vieillir » TASDA qui co-organise avec le Département de l'Isère, la Chambre de Métiers et de l'Artisanat, AG2RLa Mondiale, Minalogic et l'IRTNanoelec, la 4^{ème} édition de la Journée SilverEco sur le thème : « La maison de demain : quelles briques innovantes ? ». Et je remercie chaleureusement sa directrice Véronique Chirié de m'avoir invité à m'exprimer et à réfléchir avec vous. Je suis un chercheur en philosophie, spécialisé en éthique et en théorie politique et mes sujets d'étude concernent les innovations (technologiques, organisationnelles et sociales). C'est dans cette perspective que j'assume la direction scientifique de la chaire de recherche « éthique & IA » au sein de l'institut 3IA grenoblois MIAI¹. L'IA tend en effet aujourd'hui à devenir le système technique dominant, un système qui, *via* le numérique ce formidable amplificateur d'innovations, s'étend de la conception des algorithmes aux usages les plus variés.

A travers son implémentation sociale, l'IA engendre une situation de disruption et un danger de dépression : l'importance des éthiques de l'IA

A propos des usages sociaux et des pratiques humaines modifiées par les innovations dont l'IA est porteuse, la situation apparaît tellement nouvelle qu'elle peut légitimement sembler bouleversante ou révolutionnaire. Elle correspond à ce que le philosophe Bernard Stiegler identifiait comme la « disruption », à savoir, dans ses termes, un risque pour les contemporains de « devenir fous » (Stiegler 2016). Oui, devenir fou : les sociétés développées qui adoptent le paradigme de l'innovation que je qualifie pour ma part comme naturellement « sauvage » (Ménissier 2021) disposent de remarquables technologies, qui rendent tout ou presque possible, et connaissent également des usages d'une très grande inventivité. Oui, mais pour quel sens commun et partagé ? Cela, personne ne le sait exactement, ou plus exactement : la réponse à cette question apparaît fortement et durablement indéterminée. C'est en fonction

¹ Ce texte est le fruit du travail scientifique qui est mené dans le cadre de la chaire « éthique & IA » soutenue par l'institut pluridisciplinaire en intelligence artificielle MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

de cette situation qu'il existe peut-être bien en effet aujourd'hui un risque de « devenir fou » : en d'autres termes, la puissance de reconfiguration de la réalité propre aux technologies contemporaines et l'efficacité de l'IA n'ont aujourd'hui d'égaux que la faiblesse de symbolisation de la vie algorithmiquement assistée. Si les ingénieurs et les informaticiens ont bien œuvré, consacrant le rêve de la modernité pour la science et la technologie, pour l'instant les philosophes ont pris du retard dans la tâche qui est leur, à savoir trouver le sens véritablement commun des pratiques humaines à partir de la considération de limites clairement assignées à l'activité humaine en fonction de concepts éclairants.

Devant cette situation, je défends l'idée qu'il est nécessaire de concevoir une nouvelle philosophie des techniques et des technologies – à commencer par une philosophie de l'IA, car la philosophie du projet contemporain, qui est celui de la « computation potentiellement intégrale », fait cruellement défaut. Il convient d'élaborer une nouvelle philosophie de la technique capable de fonder une éthique/théorie politique qui soit solide et efficace. C'est-à-dire susceptible de délivrer non pas des jolies valeurs et des belles règles (cela, n'importe quelle charte – et nous avons des chartes éthiques, et par dizaines ! – apparaît capable d'en énoncer), mais des règles d'action sensées.

L'éthique complète ou réelle de l'IA gagne à se nourrir des cas d'usage offerts par l'implémentation des algorithmes dans la vie réelle des usagers. A ce titre, ce que je nomme l'éthique des usages de l'IA (qu'on pourrait également nommer *UX AI ethics*) complète les ambitions de l'éthique algorithmique (souvent nommée *Computer ethics*), de l'éthique artificielle (*Artificial ethics*) et de l'éthique du numérique (*Digital ethics*) : les quatre éthiques « régionales » ou particulières de l'éthique de l'IA ne sont pas animées par les mêmes acteurs du monde de l'IA et ne couvrent pas les mêmes angles de vue sur le même système technique mais œuvrent ensemble afin de rendre le nouveau monde davantage sensé pour une vie humaine. Et la tâche est immense, puisque les cas d'usage sont légions, le processus d'implémentation des technologies d'IA étant largement en cours de réalisation dans tous les secteurs d'activité importants pour l'existence humaine : énergie, défense/sécurité, mobilité, échanges financiers et commerciaux, santé, etc.

Je voudrais souligner que face aux cas d'usage, ce n'est pas simplement un problème d'acceptabilité qui se pose alors, par rapport à des technologies informatiques, numériques et robotiques qui seraient difficiles à faire accepter. Personnellement je récusé par principe la thématique de l'acceptabilité. En effet, elle me semble de nature à occulter le problème qui se pose aujourd'hui. Le problème ne consiste en effet nullement de « rendre socialement acceptables » les nouvelles technologies, en trouvant pour elles le bon design puis en adoptant le bon marketing. Avec des innovations très disruptives, ce serait comme « mettre sous le tapis » de nos usages éculés les problématiques totalement renouvelantes que nous permettent de vivre les nouvelles technologies. Ces dernières suscitent toutefois un véritable vertige, elles stimulent même le vertige du possible, jusqu'au bord de la dépression – ce qui est tout à fait normal : tout usager des technologies d'IA se trouve en risque de dépression car ces dernières tendent à lui faire perdre ses points de repère : par exemple, l'efficacité des systèmes qui sont déjà à l'œuvre dans les pratiques sociales menacent les emplois salariés basés sur des métiers patrimonialisés dans des traditions qui jusqu'alors faisaient sens. Y contribuent également toutes les nouvelles situations d'agentivité assistée par des technologies qui ébranlent le couple bien stabilisé constitué par la liberté et la responsabilité. Les limites mêmes de la vie et la mort sont devenues floues car il semble que le rêve transhumaniste soit, dans ses conditions mêmes de possibilité,

déjà devenu une réalité, en tout cas pour les populations aisées : il est possible de faire reculer les limites naturelles de la vie humaine de façon considérable. Mais toutes ces observations ne font que suggérer que l'époque est toute en renouvellements, et qu'il convient peut-être même de la considérer comme une nouvelle « Renaissance ». Une des tâches de la philosophie consiste aujourd'hui à élaborer la philosophie appliquée, la sagesse originale dont cette nouvelle Renaissance a besoin².

Dans ce contexte, qu'y a-t-il à dire à propos de la relation entre « IA, éthique et vieillissement » en regard des « briques innovantes » de la maison de demain ?

Aristote et l'IA, ou la recherche de ce qu'est une vie « véritablement humaine »

Il pourrait sembler étrange voire franchement décalé de se fonder sur la réflexion d'un philosophe grec. Pour approcher la sagesse dont nous avons besoin aujourd'hui en matière d'usages de l'IA, je vais en effet me référer à Aristote, le natif de Stagire en Chalcidique et le disciple de Platon, qui a vécu au quatrième siècle avant notre ère. En fait, les spécialistes de son œuvre pourraient justifier le rapprochement avec l'IA, certes sur un mode paradoxal : dans un texte célèbre de son traité *Les Politiques*, il est question d'objets techniques mobiles dont le Stagirite imagine qu'un jour ils pourraient être auto-mobiles, automatiques ou autonomes, et il évoque à ce propos les navettes d'un métier à tisser et les plectres des cithares, qui fonctionneraient tous seuls afin de rendre le service qu'on attend d'eux. Mais alors, ajoute-t-il, ces artefacts, œuvres des humains, se trouveraient enchantés comme on le dit des statues de Dédale ou des trépieds d'Héphaïstos dans la mythologie. Ce qui, conclut-il, est non seulement peu probable mais surtout rendrait caduque une institution sociale comme l'esclavage, ce qui paraît à l'auteur une vue douteuse (*Politiques*, I, 4, 1253 b).

Même si Aristote, en dépit de son imagination, ne croyait pas à ce que l'IA est aujourd'hui effectivement capable de réaliser (à savoir, rendre les machines non seulement automatiques, mais presque effectivement autonomes), je veux souligner un point de sa pensée qui me paraît utile pour notre réflexion d'aujourd'hui. A travers une œuvre gigantesque, pas uniquement philosophique mais également scientifique, ce grand savant s'est évertué à répondre à la question « qu'est-ce que, pour l'espèce humaine, vivre une vie conforme à ce que commande notre espèce, à savoir, qu'est-ce qu'une véritablement humaine ? ». Cette question c'est celle-là même qu'ont reprise les sciences humaines et sociales, toutes disciplines confondues, depuis Emile Durkheim à la fin du XIXème siècle (à la différence que pour leur part, elles ne recherchent pas à produire une forme de sagesse appropriée aux réponses qu'elles posent).

La réponse apportée par le Stagirite à la question qu'il pose est invariablement la même, sous quelque angle qu'il étudie la question (il l'aborde, tour à tour, en spécialiste de cosmologie, de sciences naturelles, de métaphysique, de psychologie des émotions, de théorie esthétique des tragédies, d'éthique ou de science politique) : une vie humaine véritablement humaine c'est une vie relationnelle, ou reliée. Relationnelle ou reliée avec les autres humains, bien entendu, mais également avec les autres vivants, avec les autres êtres naturels. Or, pour Aristote, qui sont ces autres êtres naturels ?

² L'approche de la technologie et des nouveaux usages auxquels elle donne lieu en termes de sagesse n'est pas une tendance dominante de la pensée philosophique de la technique contemporaine bien qu'on la voit aujourd'hui régulièrement resurgir (voir Puech 2008, et le programme TESaCo : <https://www.tesaco.fr/>)

Dans son ouvrage intitulé *Physique*, Aristote donne une définition très puissante de la *technê*, terme qui englobe les arts techniques, savoir, savoir-faire et savoir-être de l'action matérielle efficace à travers les outils, ces artefacts inventés de manière contrainte à cause de la faiblesse constitutive de la condition humaine. La *technê*, si on la saisit d'un point de vue général, écrit Aristote, « ou bien imite la nature ou bien effectue ce qu'elle est dans l'impossibilité d'accomplir » (*Physique*, II, 8, 199 a 15-17). Cette définition est littéralement bouleversante. En effet, elle prend le contrepied de l'artificialisme spontané, souvent exprimé par les gens qui n'arrivent pas à concevoir la relation qu'ils entretiennent avec les artefacts (instruments et systèmes techniques) qu'ils utilisent pourtant, et qui de ce fait, n'ont *pas vraiment de relation* avec eux, car ils n'ont pas conscience de la profondeur de leurs usages. Bouleversante cette phrase, puisqu'elle invite à penser que la pratique techniquement assistée, et au-delà le dispositif technique comme résultat de la *technê*, poursuivent, complètent et achèvent la nature. Du point de vue aristotélicien, l'agir technique, l'outil, la machine, le dispositif socio-technique, tous ces éléments qui relèvent de la production d'artefacts peuvent également être qualifiés de *naturels*. Il s'agit en effet pour l'espèce humaine d'activités ou de résultats de l'activité qui correspondent à sa nature propre. On voit donc qu'une éthique de la technique se déduit de l'ergonomie elle-même. En effet, normalement il n'y a nul besoin de penser une acceptabilité sociale en plus, cela serait même mauvais signe car cela signifierait qu'il est nécessaire d'user de ruses, celles du marketing, pour « survendre » quelque chose qui de lui-même peine à s'imposer dans et pour nos usages.

Je pense sincèrement la réflexion du philosophe de Stagire est non seulement très intéressante et importante pour penser l'IA dont nous avons besoin au moment de concevoir le sens philosophique et éthique de la « maison de demain », capable d'abriter les vulnérabilités du grand âge.

A la lumière de cette sagesse, il convient en effet de repenser les environnements de l'humain et de redéfinir l'autonomie

De par sa nature sociale, l'humain se fonde sur différentes sortes de relations, autrement dit il évolue dans des environnements variés, tous très importants pour le maintien de qu'il a d'humain, sans parler de la réalisation de lui-même. S'il est bien entendu impossible de penser ici tous les types d'environnements qui sont, pour Aristote, nécessaires à notre espèce, j'en aborderai rapidement deux qui sont pertinents pour la journée d'aujourd'hui.

D'abord l'environnement social. Pour nous, contemporain, cet environnement est fondamentalement urbain, et le sera sans doute encore plus demain. On entend actuellement dire qu'en 2050 au moins 70 % de la population mondiale vivra dans des villes. L'environnement social sera donc, d'après cette estimation, majoritairement urbain. Avec les problèmes que cela pose : d'ores et déjà, ou actuellement, ou jusqu'ici, les villes consomment davantage qu'elles ne produisent ce qu'elles consomment. Ces villes de demain, qu'on nous prédit intelligentes, ne le seront qu'à proportion de la qualité environnementale de l'interaction sociale. Dans les termes aristotéliciens, la cité heureuse est autarcique, et même, plus précisément, « autarcique en vue de la vie bonne » (*autarkes pros to eu dzèn, Les Politiques*, 1326 b 8-9). Il est permis de penser que les systèmes d'IA permettent d'ores et déjà de redonner un sens à la notion d'autarcie. En ce sens qu'on pourrait écrire qu'*est autarcique ce qui est durable et efficace pour des services sociaux inclusifs*. On le voit, ici, la technologie ne constitue pas un horizon en elle-même, elle ne représente pas non plus un obstacle, mais un moyen

d'atteindre ce genre de but. Encore faut-il doter la sagesse des technologies contemporaines d'un fort souci environnemental, ou même de constituer ce souci comme un des principes du développement technologique (par exemple en imposant les principes de l'écoconception, qui consistent à intégrer la dimension écologique dès la conception et le développement de produits).

Ensuite l'environnement domestique. A savoir, la « maison » (en grec *oikia*, donc à la base de toute véritable « éco-logie »), dont nous cherchons aujourd'hui à préciser les modalités grâce à l'aide conjointe des acteurs du soin pour les personnes vulnérables, des concepteurs de la technologie numérique et d'IA et des pouvoirs publics. Cet environnement est lui-même technologique, mais n'est pas réductible à l'ensemble des outils et des machines qu'il abrite. A ce propos Aristote délivre les éléments d'une définition de l'autonomie qui pourrait sembler révolutionnaire. Tellement surprenante en tout cas qu'elle conduit à la formulation d'un paradoxe : *être autonome, c'est être relié (aux autres, au système technique, à l'environnement)*. Enfin, ce paradoxe n'est qu'apparent, il repose sur l'oubli de la vraie « nature » de l'humain au profit d'une fausse idée d'autonomie, liée à la définition moderne de l'autonomie (proposée par les auteurs de l'individualisme moderne, tels Descartes et Locke). Cette illusion se nourrit d'une confusion, celle qui existe entre autonomie et toute-puissance. Le rêve de la modernité (qui trouve ses limites en regard de la crise environnementale) c'est de faire croire que la science et la technologie peuvent rendre l'humain tout-puissant.

Cette définition de l'autonomie comme capacité à être judicieusement « relié » aux autres et aux technologies permet de penser une distinction qui me paraît importante : dans la perspective d'une vie qui entend conserver un sens humain, ce qu'on peut attendre des services aux personnes vulnérables n'est pas de l'augmentation, mais de la stimulation. La conséquence en matière d'usage des technologies d'IA (si puissantes soient-elles) tient de ce point de vue à une forme de limite, sans aucun doute difficile à appréhender dans l'absolu : la frontière est sans aucun doute ténue entre l'augmentation et la stimulation, à savoir, entre l'excès technologique qui fait perdre de vue la condition humaine et la « juste mesure » de l'assistance technologique. Mais elle est capitale, ou cardinale en ce qu'elle constitue une sorte de point de repère pour le bon usage des technologies d'IA en matière de soin, de santé et d'autonomie.

On le comprend très bien en regard des enjeux de vie privée qui apparaissent, dans des appartements connectés où toute activité est potentiellement « monitorable » et mesurable. La question appelée par la distinction entre augmentation et stimulation est celle de la préservation de l'intimité : dans le vieillissement, cet aspect est sans doute le plus sensible. C'est même le point focal : faut-il tellement monitorer, surveiller et appareiller les personnes vulnérables qu'elles n'ont plus d'intimité ? L'IA qui serait véritablement « de confiance » au sens philosophique, ce serait le système technologique qui ne cherche pas à augmenter les personnes, mais à les maintenir dans ce qu'elles savent et veulent faire, qui permettrait de stimuler ces personnes pour qu'elles soient en capacité de réaliser cela, mais qui intégrerait également des fonctions pour ne pas les sur-solliciter ni les épier. Il me semble que du point de vue de la sagesse des technologies, la recherche d'une « vie accomplie », telle que l'entend le philosophe François Galichet dans sa puissante réflexion sur la fin de vie, peut parfaitement se fonder sur cette distinction (Galichet 2020).

Pour une approche éthique de l'IA en vue de son implémentation : le co-design des solutions et des valeurs pour réinventer le commun

Je voudrais enfin souligner le fait qu'aborder les questions technologiques par le biais de l'éthique, c'est faire de la réflexion prospective par le biais des démarches d'innovation. A cet égard, la primauté de l'usage est consacrée comme la meilleure façon de pratiquer de la recherche technologique. Ce qui peut sans doute sembler excessif à une approche purement technologique, mais ne saurait surprendre si l'on observe l'importance des démarches de co-conception au sein de Living Lab où les solutions technologiques sont « examinées par la pratique » et de ce fait sans cesse polies (ou rejetées) par les usagers.

Au-delà de ce fait, aujourd'hui bien établi dans les démarches d'innovation, il convient de souligner l'importance de l'approche multi-acteurs pour une véritable éthique, c'est-à-dire pour la construction d'un sens commun à partir de valeurs. Je dis cela parce qu'il me semble que la démarche dite de « l'habitat inclusif », qui va notamment vous réunir aujourd'hui, constitue une voie qui me semble intéressante et bien orientée.

Il s'agit en effet de trouver par des pratiques réglées des usages permettant un co-design qui soit un design du commun. Cela permet d'imaginer que l'IA puisse elle-même devenir un instrument capable de produire du commun, et qu'elle soit apte à nourrir des véritables projets d'autonomie pour les organisations publiques et privées.

Bibliographie

- Aristote, 1949 : *Physique*, trad. H. Carteron, Paris, Les Belles Lettres.
- Aristote, 1990 : *Les Politiques*, trad. P. Pellegrin, Paris, Flammarion.
- Galichet (François), 2020 : *Qu'est-ce qu'une vie accomplie ?*, Paris, Odile Jacob.
- Ménessier (Thierry), 2021 : *Innovations. Une enquête philosophique*, Paris, Hermann.
- Puech (Michel), 2008 : *Homo sapiens technologicus. Philosophie de la technologie contemporaine, philosophie de la sagesse contemporaine*, Paris, Le Pommier.
- Stiegler (Bernard), 2016 : *Dans la disruption, Comment ne pas devenir fou ?*, Paris, Les Liens qui Libèrent.