



HAL
open science

“ L’intelligence artificielle, le défi éthique ”

Thierry Ménissier

► **To cite this version:**

Thierry Ménissier. “ L’intelligence artificielle, le défi éthique ”. Les Cahiers de l’Atelier Arts Sciences, 2022, pp.24-27. halshs-03672104

HAL Id: halshs-03672104

<https://shs.hal.science/halshs-03672104>

Submitted on 19 May 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.

« L'intelligence artificielle, le défi éthique »

Thierry Ménissier

Paru in : *Les Cahiers de l'Atelier Arts Sciences*, n°12 2022 : *Intelligence artificielle : rencontres entre artistes et scientifiques*, Résidence 2019-2021, p. 24-27.

L'IA pose un véritable défi éthique. Lorsqu'on comprend en quoi il consiste, on saisit l'importance qu'il y a à mobiliser des artistes afin de créer les conditions d'une exploration du nouveau monde technologique qui renoue avec l'aventure humaine.

C'est la conjonction de plusieurs faits qui engendre le défi éthique posé par l'IA. D'abord, la condition socio-technique des pays développés obéit depuis un certain temps à ce que Bernard Stiegler appelait une situation de « disruption » (Stiegler 2016) : ces sociétés disposent de technologies à profusion et des usages aussi variés qu'inventifs paraissent donner une forme humaine à leur emploi. Mais l'alliance de ces deux dynamiques, celle de la technologie et celle des usages, tend à corroder la structure ordinaire du monde, qui fournit pourtant à l'activité sociale son assise. Par exemple, sous l'effet de l'automatisation et de l'ubérisation, ce ne sont pas seulement des emplois salariés (c'est-à-dire des cadres sociaux) qui disparaissent, mais également des métiers, à savoir, les cadres de la culture technique et du savoir humain.

Et cela, ensuite, pour quel sens commun et partagé ? Nul ne peut le dire avec exactitude et même pas sûr qu'il en existe un. Le politique – en tout cas tel qu'il est traditionnellement entendu sous la forme des partis politiques nationaux ou européens et des Etats représentatifs – peine à encadrer les acteurs du monde industriel. Tandis que, du point de vue social, des formes originales de domination ont déjà émergé grâce aux nouveaux outils (Casilli 2019 l'a montré de manière suggestive avec la notion de « travail du clic »).

Une philosophie de la technique fait littéralement défaut, qui serait capable tant d'éveiller les consciences au nouveau risque d'obsolescence de l'humain (Anders 1956), que de contrer les nouvelles formes de « bluff technologique » (Ellul 1988). L'effort de repenser la situation de l'humain en regard de ses nouveaux outils est d'ailleurs rendu impératif par le haut niveau actuel de fiabilité de l'expertise algorithmique. En effet, la tentation du recours à l'IA engendre une situation de confiance passive dans l'autorité des machines, préoccupante : la décision assistée par IA, assujettie au régime des *big data*, tend à devenir la norme au sein des sociétés algorithmiquement assistées dans tous les ordres humains : Justice, sécurité, gestion du risque et assurances, etc. (Ménissier 2021).

Hannah Arendt décrivait la situation engendrée par le totalitarisme du XX^{ème} siècle comme celle d'un grand désarroi critique : « La terrible originalité du totalitarisme ne tient pas au fait qu'une « idée » nouvelle soit venue au monde, mais à ce que les actions mêmes qu'elle a inspirées constituent une rupture par rapport à toutes nos traditions : ces actions ont manifestement pulvérisé nos catégories politiques, ainsi que nos critères de jugement moral. » (« Compréhension et politique », 1953). Par comparaison, l'IA semble en voie de « pulvériser » nos critères de jugement moraux. Ou si elle ne les réduit pas complètement en poudre, elle les met terriblement à mal. Dans le mode de vie algorithmiquement assisté, dans le monde de l'IoT

et des *smart cities*, que reste-t-il en effet des concepts fondateurs de la liberté humaine ? Comment concevoir désormais l'agentivité (à savoir, le fait d'être l'agent de son action) ? Que devient l'imputation causale, ce ressort de la responsabilité éthique et juridique ? Quelle valeur pour l'intuition dans le jugement réflexif, qui était jusqu'à maintenant comme une création continuée, jalonnée par des coups d'éclat et des improvisations ? Enfin quelle place pour les décisions personnelles engageantes qui ne seront pas déterminées par des données de calcul ?

Une réponse de type éthique peut-elle, dans ce contexte troublé, rouvrir des perspectives pour une existence humaine satisfaisante du point de vue de sa propre lucidité ? La dimension éthique suffit-elle pour le prix à payer lorsque cette existence se trouve littéralement livrée aux technologies du système industriel de l'innovation, c'est-à-dire aux séductions de « l'innovation sauvage » (Ménissier 2021) ? Tout dépend à vrai dire ce qu'on entend par « éthique ». Ainsi que l'a également expliqué Arendt, s'efforcer d'avoir une posture éthique revient à assumer de penser sa propre action, à la fois de manière réflexive et critique, et dans des contextes techniques, sociaux et politiques concrets. Or, la fameuse thèse de *Eichmann à Jérusalem* (1963), selon laquelle il est toujours parfaitement possible aux humains (dans des conditions où leur activité est bureaucratisée) d'agir sans penser véritablement leur action, prend, dans le contexte de l'essor généralisé de l'IA, un sens nouveau. Le concept de « banalité du mal » promue par la philosophe signifie, quel que soit le degré du mal commis, qu'une action mécanisée ou administrative effectuée dans l'inconscience des conséquences qu'elle peut engendrer, peut conduire son auteur à pactiser avec la barbarie. La recherche des conditions d'une éthique de l'IA permet-elle d'éviter les méfaits de l'algorithmisation du monde, qui est potentiellement une nouvelle forme de l'emprise bureaucratique ?

Revendiquer une approche éthique revient à assumer l'ambition d'être normatif. Certes, la normativité s'exprime de manière plurielle, et il existe plusieurs sortes de discours normatifs : religions, déontologies de profession, constitutions civiles et loi politiques, normes et règlements des organisations privées, etc. Mais l'éthique philosophique n'entend pour sa part admettre comme valables que les thèses qui se fondent sur des valeurs ultimes du bien et du juste, et les règles qu'elle institue se fonde sur des principes d'action toujours susceptibles d'être discutés par la réflexion et par le débat. C'est ainsi que l'éthique philosophique provoque pour l'existence humaine une forte et salutaire stimulation. Bien entendu, elle-même est plurielle, car elle reconnaît une variété de principes de ce genre. En voici quatre, dont on pourrait dire qu'ils regroupent l'éthique philosophique en quatre « familles » : l'éthique déontique repose sur l'expérience du devoir individuel (Kant, Dostoïevski), capable de conduire l'humain dans de véritables odyssées de sa conscience ; l'utilitarisme, sur la construction d'un raisonnement calculé en fonction de la mise en relation de la fin qu'on poursuit proportionnée aux moyens dont on dispose ; l'arétisme (Aristote) encourage quant à lui les vertus, ces dispositions innées qui se trouvent valorisées diversement par les cultures, dans la perspective du perfectionnisme moral mais aussi d'une vie sociale et politique heureuse ; enfin l'axiologisme revendique la primauté d'une valeur admise comme supérieure, ce qui redimensionne l'existence humaine individuelle : l'ordre cosmique pour les stoïciens, la République ou la nation du point de vue de la « liberté des Anciens » (Constant 1818), ou la nature pour les divers écologistes.

Un des points fondamentaux d'une éthique capable de relever le défi de l'IA consiste notamment à dépasser l'utilitarisme dans lequel s'enferme souvent le raisonnement algorithmique. La philosophe britannique Elizabeth Anscombe avait attaqué l'utilitarisme au motif qu'il délaissait la poursuite du bien pour s'attacher uniquement à une forme de

raisonnement appauvri, qu'elle proposait de désigner par le terme « conséquentialisme » (Anscombe 1958). Le succès récent du « dilemme du tramway » appliqué aux problèmes du véhicule autonome inquiète légitimement quant à l'application généralisée de ce genre d'argumentation : dans cette forme de raisonnement ultra-simplifié, qui cherche des réponses immédiates à des problèmes qui ne sont pas observés dans leurs véritables enjeux, aucune option véritablement éthique ne se dessine. Le véhicule autonome circule-t-il réellement sur des rails aussi rigides que ceux de l'antique tramway ? Et puisque le dilemme propose deux partis également létaux (le véhicule devant fatalement renverser tel ou tel genre de personnes), on n'a pas assez relevé que tout se passe ici comme dans le meilleur humour anglais ! Il apparaît au contraire nécessaire de mobiliser l'imagination tous azimuts, et de se demander quelle est la valeur des nouveaux outils technologiques en termes de stimulation de l'expérience humaine *lato sensu*. Que valent par exemple les dispositifs algorithmiques en matière d'odyssée de la conscience ? Ou en termes d'expérience de perfectionnisme moral ?

Pour répondre à de telles questions, les artistes ont un rôle majeur à jouer. En effet, leur manière de détourner le sens des dispositifs technologiques et numériques, de les concevoir de manière originale, surprenante et parfois scandaleuse, permet de faire sortir l'algorithmique et la robotique des cadres trop étroits dans lesquels la science informatique les enferme. Or, la nouvelle condition socio-technique, certes confuse, recèle de tant de possibles inexplorés ! Ceux-ci ne seront pas non plus explorés par la philosophie traditionnelle, elle-même trop souvent engoncée dans la scholastique des grands problèmes hérités. Ou, s'ils le sont, c'est imparfaitement, de manière trop mentale. Il importe aujourd'hui de ressentir des émotions. Par exemple, celle de l'enthousiasme qu'il y a à pouvoir agir dans un monde demain peuplé d'êtres artificiels nouveaux, qui ne sont pas une autre race puisqu'ils nous reflètent, et à qui nous devons apprendre autant de choses qu'ils ont à nous en apprendre. Ou encore celle du trouble sous l'illusion de sécurité produite par le système de l'IA. Ou enfin celle du courage, celui d'affronter la tentation de demander à ce système des solutions toutes faites pour nos vies complexes, ou même, face au solutionnisme technologique, celui de dénoncer l'illusion d'une réassurance machinique intégrale vis-à-vis de l'angoisse humaine.

Il importe donc aussi bien d'en appeler à une intelligence sensible de l'IA que de donner à l'imagination un rôle important dans la démarche éthique (Pierron 2012). Aussi, demandons aux artistes de jouer avec les technologies d'IA, afin de nous « entraîner à l'imagination participative » (Nussbaum 1996) en vue d'expérimenter les situations de déséquilibre relationnel qui font ressentir la plénitude bizarre de notre condition, espèce animale au milieu des autres êtres naturels mais désormais aussi confrontée à des êtres artificiels. Si elle crée effectivement des artifices, la technologie est loin d'être artificieuse ou pernicieuse, car, ainsi qu'Aristote l'avait dit en son temps de l'art technique (*technê*), elle exprime la nature humaine et même toute la nature, car « elle effectue » ce que cette dernière [*phusis*] « est dans l'impossibilité d'accomplir » (*Physique*, 199 a). Qui sommes-nous vraiment, nous qui créons ces technologies bouleversantes qui nous façonnent et nous révèlent ?

PRÉSENTATION DE THIERRY MENISSIER

Thierry Ménissier est Professeur de philosophie politique à l'Université Grenoble Alpes, chercheur dans l'Institut de Philosophie de Grenoble (IPhiG) et responsable de la chaire « éthique & IA » au sein de l'Institut pluridisciplinaire d'intelligence artificielle (MIAI) de

Grenoble. Ses recherches portent sur la transformation des principes de l'éthique publique et sur le sens de l'innovation technologique et sociale.

Dernières publications:

- *Innovations. Une enquête philosophique*, Paris, Éditions Hermann, 2021.
- « Les dispositifs de reconnaissance faciale, un défi pour l'éthique de l'IA », *Klesis. Revue philosophique*, n°49, 2021 [en ligne].
- « Un « moment machiavélien » pour l'Intelligence Artificielle. La Déclaration de Montréal pour un développement responsable de l'IA », *Raisons Politiques*, 2020/1 (n°77), p. 67-81.

Institut de Philosophie de Grenoble & chaire « éthique & IA » MIAI
Université Grenoble Alpes
Bâtiment ARSH, BP 47, Domaine Universitaire, 1281, avenue Centrale
38400 Saint-Martin d'Hères
Thierry.menissier@univ-grenoble-alpes.fr