



**HAL**  
open science

## Éthique et intelligence artificielle

Jean-Gabriel Ganascia

► **To cite this version:**

Jean-Gabriel Ganascia. Éthique et intelligence artificielle. ENA Hors les murs, magazine des anciens élèves de l'ENA, 2019. hal-03183627

**HAL Id: hal-03183627**

**<https://hal.science/hal-03183627>**

Submitted on 28 Mar 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.

# Éthique et intelligence artificielle

---

*Jean-Gabriel Ganascia*  
*Professeur Sorbonne Université*  
*Président du comité d'éthique du CNRS (COMETS)*

## **L'heuristique de la peur est-elle pertinente ?**

Dans son œuvre majeure intitulée *Le principe responsabilité*, l'un des plus grands philosophes de l'éthique du vingtième siècle, Hans Jonas, nous engage à adopter une *heuristique de la peur* qui recommande de s'inquiéter des risques que font peser les développements scientifiques et technologiques actuels sur l'humanité future, sur son existence et sur sa liberté. Cette injonction morale a reçu, et continue de recevoir, un écho considérable depuis la fin du vingtième siècle, surtout au regard des appréhensions suscitées par les menaces sur l'environnement, par exemple par la couche d'ozone ou le réchauffement climatique, et sur la santé publique à la suite en particulier de l'affaire du sang contaminé et de la crise de la vache folle. Elle donna, il y a déjà un quart de siècle, naissance au *principe de précaution* qui prescrit, même en l'absence de preuve scientifique démontrant avec précision l'existence et l'ampleur d'un danger, de prévenir des risques qui mettraient en péril l'existence humaine.

Aujourd'hui, des sommités, comme l'astrophysicien Stephen Hawking, le prix Nobel de Physique Frank Wilczek, l'informaticien Stuart Russel ou l'homme d'affaire Elon Musk nous mettent en garde. Selon eux, l'intelligence artificielle ferait peser un risque existentiel à l'humanité. Comment douter de la parole d'autorités aussi éminentes et rester sourds à leurs avertissements ? Nous devrions donc condamner, en vertu de l'heuristique de la peur de Hans Jonas, et de sa traduction juridique, le principe de précaution, les développements actuels de l'intelligence artificielle qui laissent entrevoir de si terrifiantes perspectives.

Pourtant Hans Jonas lui-même, dans un entretien avec Norbert Lossau paru dans *Die Welt*, le 29 novembre 1991<sup>1</sup>, affirme qu'il n'y a pas lieu de craindre qu'un objet technique acquiert conscience et autonomie : selon lui, la pensée consciente diffère radicalement des séquences

---

<sup>1</sup> Hans Jonas, « Les machines ne pourront jamais avoir une conscience » in *Une éthique pour la nature*, Arthaud poche, Flammarion, 2017.

d'opérations matérielles portant sur des signes qu'exécutent les ordinateurs et qu'une conscience perçoit de façon sensible avant de les traduire comme l'équivalent de pensées. L'intelligence artificielle simule, au moyen de calculs, des opérations de l'esprit qui ne prennent sens que pour le sujet qui les interprète. L'heuristique de la peur n'apparaît donc pas pertinente ici, car la crainte de voir surgir de la machine une volonté qui s'opposerait à la nôtre n'est pas fondée ; elle n'admet aucune justification tangible et ne relève que du fantasme.

## Réontologisation

Dans le même entretien qu'il eut en 1991 avec Norbert Lossau, Hans Jonas explique que les machines ne pourront jamais ôter aux femmes et aux hommes la responsabilité des décisions qu'ils prennent, mais que, « si l'introduction de systèmes informatiques équivalents devait avoir pour conséquence que l'importance capitale de l'individu dans sa singularité soit minée au profit d'une machinerie sociale [...], cela serait grave ». Or, de nos jours, la part que prennent le numérique et l'intelligence artificielle tant dans les échanges que dans les décisions devient si prépondérante qu'il arrive que le verdict des machines l'emporte sur le nôtre et que nous nous trouvions alors confronté à des situations analogues à celles que présentait Hans Jonas. Ainsi, arrive-t-il que ces sibylles de l'hyper-modernité que sont les ordinateurs rendent d'obscurs oracles auxquelles on tend à soumettre les femmes et les hommes de notre temps au nom d'une pseudo rationalisation de l'action. L'ouvrage de Cathy O'Neil, *Weapons of Math Destruction* en fournit de nombreuses illustrations dans les champs de l'éducation, des assurances, de la justice, de l'administration etc. Il montre l'effet prédateur d'une utilisation sans discernement d'indicateurs acquis par apprentissage machine sur de grandes masses de données (*Big Data*). À cet égard, le vote, en 2015, d'une résolution du parlement européen demandant, à terme, de créer une personnalité juridique des robots paraît révélateur d'une telle abdication.

Plus généralement, les notions multiséculaires qui font la trame du tissu social évoluent avec le numérique et l'intelligence artificielle. Ainsi en va-t-il de l'amitié qui se redéfinit doublement avec les réseaux sociaux, d'une part en ce que se fait jour une nouvelle notion d'amitié caractérisée par un lien virtuel sur les réseaux, d'autre part parce que l'attention à l'ami cher tend à se distendre du fait même de ce lien virtuel qui ne nécessite plus de parler pour prendre des nouvelles. De même, la réputation qui reposait, jusqu'à présent, sur l'appréciation des personnes de votre voisinage, se calcule désormais à l'aide d'algorithme en agrégeant des

informations diverses. Des bruits courent selon lesquels la Chine conduit actuellement des expérimentations visant à établir automatiquement un score de réputation pour l'ensemble des citoyens, en fonction de leur comportement repéré avec des techniques d'intelligence artificielle comme la reconnaissance faciale. Toutefois, l'occident démocratique n'est pas en reste : les institutions financières recourent aussi à des techniques d'intelligence artificielle utilisant des informations sur les comportements individuels, afin d'attribuer des prêts bancaires... D'autres concepts centraux, comme la confiance, songeons à la *Blockchain*, la monnaie, avec par exemple le *Bitcoin*, ou la souveraineté évoluent considérablement avec le numérique. Il s'ensuit que toutes les pratiques sociales, les *habitus*, les usages, à savoir ce que l'on appelle les mœurs, se modifient. Avec eux, les règles de bonne conduite, autrement dit les maximes de la morale doivent aussi se transformer. Or, celles-ci ne peuvent pas s'établir à partir de traditions ou de religions, et ce pour au moins deux raisons : d'une part, les conditions d'existence changent d'une façon inédite, ce qui fait que nous nous trouvons dans un monde qui n'a jamais existé auparavant et où, de ce fait, la tradition, c'est-à-dire la transmission et la reprise des usages du passé, n'est d'aucun secours, d'autre part, l'extension planétaire des réseaux empêche que les usages locaux et les croyances religieuses, s'imposent à tous. Il incombe donc à l'éthique de penser les fondements de règles de conduite adaptées à la société numérique et de réfléchir à leur mise en œuvre.

## Éthiques de principes

Depuis quelques années, un nombre très conséquent d'institutions de tous ordres, sociétés savantes, comme IEEE, universités, comme l'université de Montréal, clubs de réflexion, comme *l'institut du futur de la vie*, commissions gouvernementales, comme la chambre des lords en Grande-Bretagne, ou supra-gouvernementales, comme la commission européenne, réunissent des groupes de réflexion pour légiférer sur l'éthique de l'intelligence artificielle. Ceux-ci rédigent des chartes, des règles et des guides que nous n'énumérerons pas ici faute de place, mais qui se fondent tous sur l'énoncé de grands principes. Jusqu'ici, on en dénombre pas moins de 67 auxquels on doit ajouter une somme tout aussi longue de recommandations. Parmi ceux-là, beaucoup apparaissent indiscutables. Ainsi en va-t-il de la protection de l'intimité de la vie privée, de l'autonomie du sujet, de la solidarité, de la sécurité et de la vie démocratique ou de la promotion de la transparence, de la diversité, de l'inclusion et de l'équité. Toutefois, ils paraissent trop nombreux pour que les ingénieurs puissent les prendre tous en considération

lors de la conception de dispositifs innovants, au risque sinon d'assécher totalement leur créativité. De plus, ils se révèlent contradictoires : comment concilier les exigences légitimes de protection de la vie privée et de transparence, ou d'inclusion et d'équité ? Enfin, certains sont discutables comme l'adoption d'une approche résolument « centrée sur l'homme » dans le rapport<sup>2</sup> rédigé en 2019 par le groupe de 52 experts dits de « haut niveau » réunis par la commission européenne. Ce qui se donne ici comme humaniste suppose en réalité qu'il pourrait exister une autre approche « centrée sur la machine », et laisse ainsi place à des thèses transhumanistes extrêmement problématiques.

## Questions vives

En conclusion, la réflexion éthique suscitée par les développements de l'intelligence artificielle se trouve désormais placée dans une situation délicate où des exigences toutes légitimes entrent en tension les unes avec les autres. Comment opérer pour faire, et surtout faire comprendre, les compromis nécessaires ? Telle sera la tâche délicate à laquelle nous nous confronterons dans les années qui viennent. À défaut, toute tentative de réforme se heurtera à une opposition franche d'une partie de la population. On se souviendra, à cet égard, de la loi sur le renseignement de 2015 qui suscita une grande hostilité, alors qu'elle tentait de donner un cadre légal à l'action de l'État en matière de renseignement pour assurer la sécurité intérieure à l'ère des réseaux numériques.

---

<sup>2</sup> High Level Expert Group on Artificial Intelligence, *Ethical Guidelines for a Trustworthy AI*, European Commission, April 2019 (<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>)