La morale

L'éthique artificielle ou l'éthique d'après l'intelligence artificielle

Laurent Cournarie

Philopsis : Revue numérique
https://philopsis.fr

L'intelligence artificielle (*IA*)[1] est en train de bouleverser nos vies sous nos yeux, malgré nous et avec notre consentement. Ce n'est pas la première révolution. Lévi-Strauss considérait qu'il y avait eu deux révolutions majeures dans l'histoire de l'humanité : la révolution néolithique et la révolution industrielle. Il se pourrait que la révolution numérique soit la troisième du genre. Elle en possède la radicalité, mais avec pour spécificité d'être une révolution de et par l'intelligence. Révolution par l'intelligence — on est passé à une société de la

1 L'«intelligence artificielle » (*IA*) désigne la partie de l'informatique qui a pour objet la création d'un programme informatique capable de conférer à une machine dans laquelle il est implanté un comportement "intelligent". L'« éthique artificielle », ici prise pour synonyme des expressions « *robots ethics* » ou « *machine ethics* » plus couramment utilisées, est la partie de l'intelligence artificielle qui a pour objet la création d'un programme informatique capable de conférer à une machine dans laquelle il est implanté cette forme particulière de comportement intelligent qu'est un comportement moral, c'est-à-dire intégrant la polarité du bien et du mal.
Nous proposons cette expression : "l'éthique d'après l'*IA*" à la place de "l'éthique appliquée à l'*IA*" qui sert d'habitude à définir l'éthique artificielle pour suggérer que l'*IA* n'est pas seulement l'objet de l'éthique, mais aussi le sujet de l'éthique, que l'*IA* est susceptible de transformer l'éthique qui, par elle-même, n'est évidemment rien de stabilisé.

connaissance, c'est-à-dire à une société où tous les rapports sont médiatisés par des systèmes informatiques ; révolution de l'intelligence parce que l'intelligence est (dite) artificielle.

On le sait, l'*IA* suscite le mythe : soit l'utopie d'un monde où le travail pénible et répétitif sera délégué à des machines, où la maladie pourra être éradiquée, où le crime prédit, etc. — bref où le mal sera annulé ; soit l'apocalypse d'un monde ayant atteint le point de singularité où l'IA forte remplacera l'intelligence humaine (naturelle) et deviendra(it) potentiellement hostile à l'humanité — bref où le cauchemar sera devenu réalité.

L'avenir ne sera ni l'utopie d'une humanité libérée par l'*IA* de toutes ses servitudes, ni la dystopie d'une humanité asservie à l'*IA*, l'histoire étant, même et surtout technologiquement, imprévisible. Mais il n'en demeure pas moins que l'*IA* pose, au-delà des défis techniques, de multiples questions à la fois sociétales, juridiques et éthiques.

Le thème « *IA* et éthique » est très débattu et déjà rebattu. Certains peuvent considérer que l'éthique en *IA* est un thème marketing (*AI for Humanity*) et, qu'à s'en faire le spécialiste, on risque de manquer le développement scientifique et économique de et par l'*IA* : le gagnant en éthique serait le perdant en économie. L'éthique de la responsabilité, pour ainsi dire, recommanderait de surmonter la réflexion éthique sur l'*IA* pour se lancer dans le financement et la recherche en *IA* : la recherche de l'éthique en *IA* ne serait que le supplément d'âme de la recherche scientifique et industrielle. D'autres au contraire, considérant, au nom d'une sorte d'éthique de la conviction, que certaines valeurs éthiques sont essentielles à l'humanité, que l'*IA* doit augmenter l'homme, non le remplacer, s'élèvent contre toute naïveté à croire que les algorithmes sont neutres, qu'ils ne recèlent aucun biais cognitif, que leurs résultats en *deep learning* sont transparents et explicables. L'éthique ici se présente comme une intelligence critique (des dérives) de l'*IA*. Donc on s'inquiète[2] ainsi des effets éthiques de l'*IA*, en insistant sur la nécessité : (a) de protéger les données personnelles — l'éthique commence par le respect de la liberté individuelle ; (b) de lutter contre les discriminations — l'éthique repose aussi sur le principe d'égalité des personnes[3] ; (c) de préserver le pluralisme humain contre une tendance à la « clusterisation » des individus — l'éthique suppose le pouvoir pour l'individu d'être le sujet propre de sa vie propre ; (d) la nécessité de défendre le bien commun ou l'intérêt général des sociétés et peut-être de l'humanité en général contre d'une part des entreprises privées et d'autre part contre l'individualisme de l'usager et du consommateur — l'éthique est aussi une question politique.

Mais plus précisément comment l'*IA* interroge-t-elle l'éthique ou comment l'éthique peut-elle se saisir de l'*IA*, si l'on ne confond pas tout ce qui n'est pas technique avec l'éthique et l'éthique avec le sociétal ? Que fait l'*IA* à l'éthique ? L'éthique est à la mode et l'*IA* s'impose à la réflexion collective. Alors comment traiter du rapport entre éthique et *IA* ? On peut avancer trois idées simples : (1) le monde technoscientifique contemporain n'est pas moins éthique mais plus ; (2) cette croissance éthique est peut-être en même temps une crise morale ; (3) l'enjeu éthique de l'*IA* se concentre sur l'autonomie humaine de la décision.

(1) Le monde est plus éthique sous deux aspects. Il l'est d'abord « sociétalement ». L'éthique est à la mode et elle est partout : éthique des affaires, éthique des entreprises, bioéthique, éthique environnementale, éthique animale, etc. Tout est éthique ou tout socialement pose un problème éthique. Mais il l'est socialement parce qu'il l'est surtout problématiquement. Relève désormais de l'éthique ce qui jusque-là n'en avait jamais relevé. L'éthique désignait et désignait toute théorie normative de l'action : comment bien agir ou que doit-on vouloir ? — étant admis que toutes les actions ne se valent pas, que les sociétés valorisent certaines comme bonnes ou justes et d'autres comme mauvaises ou injustes. Il n'y a pas d'éthique ou de morale en deçà de cette distinction entre du bon et du mauvais et du désir ou de la volonté d'orienter

2 Cf. le rapport de la CNIL (décembre 2017) sur Les enjeux éthiques des algorithmes et de l'intelligence artificielle.

3 Entre autres exemples, le programme *IA* d'*Echo look*, un coach stylistique lancé par Amazone, avait éliminé les candidats de couleur noir, ayant déduit à partir des données fournies, que la peau claire était un critère de beauté.

l'action en fonction de ces valeurs. Ce qui a changé c'est, pour ainsi dire, le périmètre de l'éthique ou de la morale. Jusqu'à présent, la norme éthique concernait : 1. le rapport de l'homme à l'autre homme — étaient exclus de l'éthique la nature, l'animal, la machine ; 2. dans les limites du temps présent, au moins possible (devoirs envers le prochain) — étaient exclus de l'éthique les générations à venir ; 3. et plus fondamentalement, dans ce qu'on peut appeler les limites de la finitude humaine — étaient exclus de l'éthique tout ce qui touche aux fins de l'existence (naissance/mort)[4]. Or ce que nous subissions, nous sommes en passe de pouvoir le vouloir : l'impossible devient possible. *From chance to choice*. La nature et le hasard ont longtemps décidé de la naissance, de la vie et de la mort. L'extension du périmètre de l'éthique est ainsi en réalité l'augmentation (du cercle) de la responsabilité humaine. Le monde est donc plus éthique parce que nous devons collectivement normer nos actions sur la naissance, la vie et la mort qui avaient, au moins dans leur possibilité, toujours échappé au champ de la responsabilité morale.

(2) Mais en même temps le monde est peut-être moins moral. On assiste à un recul de la capacité à poser une loi catégorique ou à interdire inconditionnellement une action pragmatiquement ou techniquement possible (loi de Gabor). L'humanité vit en quelque sorte dans un autre monde éthique. L'ancien monde était fait de devoirs et d'interdits, investis et cautionnés par des institutions. Or, pour des raisons multiples[5], il est de moins en moins possible d'imposer une règle pour interdire un projet, voire seulement de prolonger un moratoire. Le « non » s'efface de l'horizon notre monde. Les intérêts économiques, la disparité des systèmes juridiques, le désir d'apprendre et de connaître davantage, la revendication des droits et de l'égalité et, tout simplement, ce qu'il est convenu d'appeler l'évolution des mœurs, sont plus puissants que les scrupules moraux ancestraux.

C'est à l'aune de ce double contexte (inflation éthique, déflation morale) qu'il faut situer l'enjeu éthique de l'*IA*. Pourtant l'*IA* pose des problèmes éthiques spécifiques. Que peut-être l'éthique « d'après » l'*IA* ? En quel sens peut-on parler d'une éthique artificielle ? Car il ne s'agit plus simplement de savoir comment l'homme doit user des machines, il s'agit de savoir comment il doit programmer les machines pour un comportement éthique intelligent. L'éthique passe de l'usage humain de la machine (et plus généralement de la technologie) à la capacité d'action autonome de la machine. Autrement dit, l'éthique artificielle ne désigne pas seulement l'éthique appliquée à l'*IA* (l'*IA* comme objet de l'éthique[6]) mais l'*IA* comme sujet de l'éthique.

(3) L'enjeu éthique de l'*IA* se concentre sur l'autonomie humaine de la décision. L'*IA* ne menace-t-elle pas le pouvoir de l'individu (mais aussi des groupes) de décider lui-même de sa vie, de maîtriser sa décision ?

Une bonne illustration de cet enjeu est le cas de l'automobile autonome. Il ne s'agit pas seulement un véhicule sans conducteur embarqué — un drone est de ce type, mais il est piloté à distance, sorte de « télé-véhicule »[7]. L'automobile autonome est un robot : c'est encore (extérieurement) une voiture, mais dotée d'une *IA*, constituée de milliers de lignes de

4 Descartes, au XVIIème siècle, écrivait que « s'il est possible de trouver quelque moyen qui rende communément les hommes plus sages et plus habiles qu'ils n'ont été jusqu'ici, je crois que c'est dans la médecine qu'on doit le chercher » (*Discours de la méthode*, VI). Mais désespérant des progrès de la médecine à laquelle il avait consacré beaucoup d'efforts, il estime finalement dans une *Lettre à Chanut* (15 juin 1646) qu'« au lieu de trouver les moyens de conserver la vie, j'en ai trouvé un autre, bien plus aisé et plus sûr, qui est de ne pas craindre la mort ». Autrement dit la maîtrise de soi, de ses désirs et de ses craintes, l'effort pour acquérir des vertus (courage, tempérance, justice…) au fondement de l'exigence éthique sont peut-être la seule réponse que l'humanité pouvait se donner tant qu'elle était impuissante à peser sur le destin de sa vie biologique.

5 Recul de l'autorité des institutions, individualisme, pluralisme des valeurs dans des sociétés de plus en plus multiculturalistes, compétition mondiale, etc.

6 L'*IA* interroge l'éthique de manière très variée, on l'a vu : opacité des processus de prise de décision (le problème de la boîte noire, Knight, 2017), la fracture entre les pays dans le développement de l'*IA*, la juste répartition de la croissance de la productivité par l'*IA*, la surveillance et la discrimination invisibles : cf. Martin Gibert, « L'éthique artificielle », *l'Encyclopédie philosophique* : http://encyclo-philo.fr/etique-artificielle-gp/.

programme informatique articulées à des capteurs multiples qui lui permettent d'agir et de réagir à un environnement en mouvement. Si agir consiste à produire un mouvement intentionnel dont l'agent contrôle en permanence l'adéquation avec le but recherché, alors une "autonomobile", comme on se propose de l'appeler, est définissable comme un « agent pratique artificiel modulaire » (*APAM*) — mais c'est également vrai d'un robot aide-soignant : (a) un agent « pratique » parce qu'elle poursuit de manière intelligente un but pratique en s'adaptant à ses objectifs (feux de signalisation, limitation de vitesse, passage de piétons, circulation…) et prenant d'elle-même des décisions pratiques appropriées (ralentir, freiner, s'arrêter, changer d'itinéraire…) ; (b) un agent pratique « artificiel » parce qu'elle est une *I A* (programme informatique qui commande des séries d'impulsions électroniques) ; (c) un agent pratique artificiel « modulaire » parce qu'elle n'est capable, contrairement à l'agent humain, que d'un spectre restreint de buts pratiques[8].

7 Pour ce qui suit, notamment sur le concept d'*APAM*, nous nous référons aux réflexions de Stéphane Chauvier dans son article « L'éthique artificielle », *l'Encyclopédie philosophique* : http://encyclo-philo.fr/ethique-artificielle-a/

8 Par exemple, *lethal autonomous weapons systems*, ou robots aides-soignants, voire robots d'assistance sexuelle cf. D. Levy, *Love and Sex with Robots*, New York, Harper Collins, 2007.