What contribution can ethical and political philosophy make to the debate on the ethics of AI?

Thierry Ménissier

Thierry Ménissier

Institut de Philosophie de Grenoble, Université Grenoble Alpes & Chaire éthique&IA, MIAI

Thierry.menissier@univ-grenoble-alpes.fr

« Quel apport pour la philosophie éthique et politique dans le débat sur l'éthique de l'IA ? », participation à la table ronde « IA et numérique : Éthique, impact sur la société et sur la science »**,** Journées Scientifiques INRIA, Montbonnot, 29 août 2024.

Titre en anglais de la contribution : « What contribution can ethical and political philosophy make to the debate on the ethics of AI? »

Argumentaire de la table ronde : L'ambivalence du numérique et de l'IA n'est plus à démontrer : à côté des bienfaits classiquement associés à la numérisation du monde, on trouve pêle-mêle capitalisme ou communisme de surveillance, biais et discriminations algorithmiques, fractures numériques, bulles informationnelles et hystérisations du débat public, sans parler de l'empreinte éco-systémique croissante des technologies numériques, exacerbée par le déploiement massif de l'IA. La table ronde s'efforcera d'apporter quelques éléments d'analyse pour comprendre cette ambivalence, la place des technologies numériques et de l'IA dans des sociétés souhaitables, ainsi que le rôle et la responsabilité des scientifiques dans l'élaboration de ces technologies.

https://jsi2024.inria.fr/programme/

Résumés de l'intervention :

FR : On comprend presque intuitivement l'apport des études juridiques et des mathématiques informatiques au débat sur l'impact de l'IA et du numérique en matière d'éthique : le droit exerce sa vocation normative, les concepteurs du numérique et de l'IA expriment leur déontologie professionnelle. L'apport de la philosophie est moins évident, il ne s'impose pas de lui-même. Dans cette intervention, nous voulons préciser quel type de philosophie peut contribuer par des apports originaux et utiles au débat sur l'évaluation éthique de l'IA et du numérique.

ENG : The contribution of legal studies and computer mathematics to the debate on the ethical impact of AI and digital technology is almost intuitive: law exercises its normative vocation, and digital and AI designers express their professional deontology. Philosophy's contribution is less obvious, and not self-evident. In this talk, we aim to clarify the type of philosophy that can make original and useful contributions to the debate on the ethical assessment of AI and digital technology.

Argumentaire de l'intervention :

Mon apport à cet échange concernera la philosophie éthique et politique de l'innovation. J'évoquerai d'abord les deux niveaux sur lesquels on attend généralement une telle entreprise théorique. Dans un second temps, je poserai comme hypothèse qu'un niveau médian permet de les associer, niveau sur lequel apparaît une tâche importante : construire des IA d'intérêt général.

Également nécessaires, les deux niveaux sont toutefois radicalement différents, le premier se voulant « réaliste », le second tourné vers la compréhension du rôle de la technologie pour une société souhaitable.

D'une part, une philosophie éthique et politique de l'innovation gagne à être « réaliste » : à l'instar des grands modèles du genre (Thucydide, 1990 ; Machiavel, 1996 ; Aron, 1965), il est

possible et fécond d'analyser l'histoire du développement de l'IA à l'aune des relations qu'entretiennent des puissances sociales (les Etats et les firmes). Dans cette optique, il est difficile de ne pas admettre que le développement de l'IA dans la société consacre le paradigme de "l'innovation sauvage" (Ménissier 2021) : s'il obéit également à une logique scientifique, il reçoit un bon accueil sociétal – et la plupart de ses financements – non pas parce qu'il est scientifiquement intéressant, mais parce qu'il est étroitement lié aux attentes de notre société, qu'on peut qualifier de capitaliste et d'industrialiste sur le plan économique. On parle ici d'une logique de rentabilité favorise l'accumulation des moyens et la concentration des richesses, dans un contexte de concurrence féroce masqué par les séductions du marché (dans le paradigme de l'innovation, c'est la logique des usages qui anime la croissance). Ainsi, la question de la régulation juridique et de l'évaluation éthique s'inscrit dans un contexte dont il serait illusoire de les couper. Et plusieurs questions se trouvent ici solutionnées, comme celle de la supposée neutralité de la science informatique. Pour se voir socialement encouragée comme elle l'est, celle-ci ne saurait être neutre.

D'autre part, ce qu'on peut attendre d'une philosophie de l'innovation est qu'elle interroge les finalités de la technologie pour la société – si possible pour une société souhaitable, ce qui qualifie la philosophie comme une discipline non pas seulement descriptive, mais normative ou évaluative. Or, l'IA n'a que peu fait l'objet d'une telle tentative, tandis que ces trente dernières années, faisant suite à la vague des penseurs des années 1970 (Ellul, Illich), plusieurs tentatives importantes ont entrepris de questionner le sens de la technologie (incluant les technologies de l'information et de la communication) pour une telle société (Feenberg, 1999 & 2002 ; Stiegler, 2018 ; Galimberti, 2023). A cet égard, si de salutaires travaux à portée critique se multiplient (O' Neil, 2016 ; Crawford, 2021, etc.), ceux à portée normative n'existent pas encore. Comment est-il possible de construire une philosophie de l'IA en vue d'une société souhaitable ? A noter toutefois qu'à de rares exceptions près, ces tentatives n'incluent pas encore la dimension environnementale ; les réflexions les plus avancées en la matière évoquent l'héritage des Lumières à l'âge du vivant (Pelluchon, 2021).

Plusieurs verrous entravent une telle tentative, et ils sont variés et à certains égards contradictoires les uns avec les autres. A ce stade, nous en identifions quatre :

- Le manque de recul sur ces technologies souvent encore au stade de sortie de laboratoire : on ne sait pas jusqu'où on peut aller en matière de calcul, situation propice à la course en avant,
- (Et paradoxalement) le caractère sauvage de l'innovation tournée vers une mise en marché hyper-rapide : il y a des usagers-consommateurs prêts à utiliser ces technologies mises à leur disposition,
- L'interaction des intérêts entre les acteurs publics et privés en vue du développement de l'IA : elle manifeste ce qu'on pourrait métaphoriquement désigner par le « désir d'IA », ou désir pour l'IA que l'on pourrait décrire ainsi : le développement exponentiel des SIA vient du fait qu'il y a consensus pour *ne pas penser*, ni la vacuité de l'existence personnelle (on peut désormais la remplir de *data*) ni les contradictions sociales (l'IA a tendance à s'implanter dans les vides sociétaux, qui sont souvent des zones sensibles pour la décision humaine : par exemple fonctions de la Justice, du travail industriel ou tertiaire – par exemple la fonction RH – optimisés : le consensus des acteurs repose sur l'idée que *l'IA passe pour être la clé d'une société optimisée parfaite*),

- La non-considération des coûts directs et indirects de ce dernier sur l'environnement ou l'absence d'effets sur les concepteurs d'IA de la crise climatique.
- Le fait que ce qu'on appelle éthique de l'IA se trouve dominé par une forme de raisonnement utilitariste ne permettant pas toujours de réellement aborder les questions éthiques dans ce qu'elles ont de radical (Ménissier, 2023).

A ce stade, nous soutenons que les tentatives de régulation juridique ou d'évaluation éthique de/sur l'IA sont en-deçà des enjeux aujourd'hui rencontrés par les individus et les sociétés.

Pour sortir du dilemme, une piste raisonnable qu'on peut actuellement proposer est de « recomposer l'intérêt général » à propos de l'IA, par exemple selon un modèle que nous avons autrefois proposé dans le cadre de la théorie normative. Que seraient des systèmes d'IA d'intérêt général ? La première étape pour répondre à cette question est de faire émerger une « généralité » pour l'intérêt porté à l'IA – tandis qu'une cacophonie d'intérêts particuliers se font entendre et s'associent pour une doctrine utilitaire sans réflexivité ni profondeur. La généralité de l'intérêt devra désormais inclure les intérêts de la nature et du vivant. Telle semble être la tâche cardinale d'une éthique politique de l'IA d'inspiration réaliste et à portée philosophique. C'est à une conversation portant sur ce sujet que sont aujourd'hui conviés tou.te.s les acteurs et actrices raisonnables réunis autour de l'IA. Il apparaît dons souhaitable, si l'on souhaite réellement parvenir à une éthique pour l'IA en responsabilisant les acteurs/actrices, de « politiser » la conversation.

Références :

Aron, Raymond (1965) : *Démocratie et totalitarisme*. Paris : Gallimard.

Crawford, Kate (2021). *Atlas of IA*. New Haven: Yale University Press.

Feenberg, Andrew (1999) : *Questioning Technology.* Londres : Routledge.

Feenberg, Andrew (2002) : *Transforming Technology. A Critical Theory Revisited*. Oxford : Oxford University Press.

Galimberti, Umberto (2023, 1ère éd. 1999) : *Psiche e techne. L'uomo nell'età della tecnica*. Milan : Feltrinelli.

Machiavel, Nicolas (2004) : *Discours sur la première décade de Tite-Live*. Traduction A. Fontana et X. Tabet. Paris : Gallimard.

Ménissier, Thierry (2009) : « Recomposer l'intérêt général. Un essai de théorie normative en réponse à la crise du républicanisme classique », *Dissensus*, Université de Liège, n°2/2009, p. 178-199.

Ménissier, Thierry (2021) : *Innovations. Une enquête philosophique*. Paris : Hermann.

Ménissier, Thierry (2023) : « Les quatre éthiques de l'intelligence artificielle », *Revue d'anthropologie des connaissances*, 17-2 | 2023.

O'Neil, Cathy (2016). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. New York : Crown Publishing Group.

Pelluchon, Corine (2021) : *Les Lumières à l'âge du vivant*. Paris : Editions du Seuil.

Stiegler, Bernard (2018) : *La Technique et le Temps. 1. La faute d'Épiméthée* (1994), *2. La désorientation* (1996), *3. Le temps du cinéma et la question du mal-être* et *Le nouveau conflit des facultés et des fonctions dans l'Anthropocène* (2001). Paris : Fayard.

Thucydide (1990) : *Histoire de la guerre du Péloponnèse*. Traduction J. de Romilly. Paris : Robert Laffont.