« IA : Que faire de la transparence technique ?»

Colloque international organisé dans le cadre de la Chaire Éthique & IA

de l'Institut MIAI à l'Université de Grenoble-Alpes

8-9 octobre 2020

Présentation de la Chaire

La chaire Éthique & IA fait partie de l'Institut MIAI Grenoble Alpes (Multidisciplinary Institute in Artificial Intelligence)[1] et se rattache à l'Institut de Philosophie de Grenoble (IPhiG)[2]. Elle vise à développer sur une durée de quatre ans (2019-2023) une connaissance de l'intelligence artificielle dans une approche philosophique qui soit en dialogue permanent avec l'informatique et la robotique, la psychologie cognitive, sociale et clinique, les sciences de l'information et de la communication ainsi que les sciences de gestion. A la croisée de la philosophie politique, de la philosophie des techniques et de l'éthique publique, la chaire vise à explorer et à comprendre les enjeux sociaux, moraux et politiques du déploiement des technologies de l'IA de manière à la fois critique et respectueuse de leurs réalités techniques.

Argument du colloque

Le déploiement de processus algorithmiques capables d'apprendre à partir de données et de participer activement à une pluralité de pratiques sociales (dépistage de cancers, évaluation du crédit, recommandation de contenus culturels, etc.) pose la question de leur « transparence ». Mais la teneur de ce terme, dont l'attrait semble irrésistible et la valeur irréfutable depuis une vingtaine d'années, est loin d'être aussi évidente qu'il n'y paraît. En effet, il semblerait qu'on soit face à une double transparence, caractérisée par une tension encore peu ou faiblement problématisée.

D'une part, les systèmes algorithmiques apprenants (*machine learning*) avec lesquels nous interagissons dans un nombre croissant de domaines sont encore largement conçus selon un idéal de transparence que l'on pourrait qualifier de *phénoménologique* (M. Wheeler ; D. Ihde ; M. Heidegger). De ce point de vue, l'objet technique qui fonctionne normalement, qui remplit sa fonction, est celui qui s'efface dans l'action, qui disparaît du champ de perception immédiat pour ouvrir des possibilités d'action amplifiées. Ainsi, nous n'avons pas conscience du crayon lorsque nous écrivons normalement mais uniquement lorsque la mine se brise ; de même les lunettes de vue ne sont perceptibles qu'à partir du moment où elles présentent une anomalie qui gêne la vue normale. On voit cette logique se transposer au niveau du design de systèmes algorithmiques qui soient ergonomiques et *user-friendly.* L'objet technique est censé être aussi discret que possible, voire invisible, c'est-à-dire que l'on doit pouvoir voir à travers lorsque nous agissons avec lui. On retrouve cette logique à l'œuvre dans le célèbre PageRank de Google qui demande explicitement aux éditeurs de contenus de ne pas prendre en compte les critères de l'algorithme dans leurs publications, de faire comme s'ils agissaient « naturellement » (D. Cardon). La même logique se rejoue sur les plateformes comme YouTube, Amazon ou Netflix dont l'action de recommandation est censée être

[1] https://miai.univ-grenoble-alpes.fr/

[2] https://iphig.univ-grenoble-alpes.fr/programmes-recherche/chaire-ethique-ia

aussi fluide et imperceptible que possible (T. Reigeluth). La question est de savoir si le critère normatif mis en avant par le paradigme de la transparence phénoménologique est réellement transférable ou même désirable lorsqu'il s'agit d'objets techniques qui ne sont pas de simples instruments pour nos sens et nos organes, mais des systèmes faisant preuve d'une certaine autonomie (M. Wheeler) et inventivité normative, en ce qu'ils transforment les pratiques sociales dans lesquelles ils sont déployés (J. Grosman et T. Reigeluth). En d'autres termes, on est en droit de se demander si l'enjeu majeur ne consiste pas plutôt, à rebours de cette première tendance, à rendre les systèmes techniques *aussi perceptibles que possible*, ou du moins à repenser la manière dont ils se présentent à nous dans l'action.

D'autre part, les critiques – provenant pour la plupart des sciences humaines et sociales ou de la théorie du droit – préconisent que les effets normatifs produits par ces systèmes (biais, discriminations, etc.) soient rendus lisibles et transparente (N. Diakopoulos ; F. Pasquale, C. Sandvig), dans un mode de gouvernement faisant la part belle à la reddition des comptes (*accountability*). Toutefois, rien ne nous dit qu'une telle exigence soit suffisante pour gouverner la complexité de ces systèmes techniques, qu'elle nous permette d'accéder à une vérité plus profonde que celle qui se joue dans les effets normatifs (bien réels au demeurant) qu'ils produisent. Le paradoxe de la plupart de ces critiques « sociales » est qu'elles reproduisent une idée de transparence des états cognitifs, mise en avant par certaines approches des sciences cognitives et neurologiques (Réf), selon laquelle il existerait une correspondance observable entre un comportement et un état cognitif cérébral. Or, lorsque les ingénieurs qui programment ces algorithmes ont eux-mêmes du mal à rendre compte précisément des raisons pour lesquelles les systèmes algorithmiques ont produit telle ou telle sortie problématique sur le plan social ou éthique, la question se pose de savoir si le fait de pouvoir « ouvrir la boîte noire » servira à quelque chose. Et quand bien même cela serait utile, voire nécessaire, il n'est pas certain que cela soit *suffisant* pour gouverner ces systèmes (J. Burrell). En effet, en tant que mode de gouvernement, la transparence pose une série de difficultés et de limites qu'il s'agit de penser : surcharge informationnelle, règne de l'expertise, mise à plat et indistinction des enjeux politiques, difficulté à cerner le public concerné par la transparence, sur-responsabilisation des ressources et compétences individuelles à décoder l'information (M. Ananny et K. Crawford ; T. Berns).

Le risque alors serait que la transparence soit finalement une exigence relativement faible ou lâche, qui s'apparente davantage à un mot d'ordre avec lequel tout le monde est, dans l'absolu, d'accord. En effet, qui s'opposerait à la transparence, quand celle-ci est devenue synonyme de bonne gouvernance, voire de démocratie ? Bien plus : comment arriver à une norme politique forte qui soit techniquement efficace et actionnable ? Dès lors, la question suivante se pose : ***que doit-on faire de la transparence ?*** Ce colloque propose de répondre à cette question à partir de la pluralité des cadres théoriques et des domaines scientifiques concernés par la question de la transparence des systèmes algorithmiques apprenants : sciences cognitives, philosophie politique et des techniques, informatique, sociologie, anthropologie, sciences de l'information et de la communication, sciences de gestion, théorie du droit, etc. Les langues du colloque sont le français et l'anglais, il est fortement conseillé aux participant.e.s qui proposent une intervention en anglais d'avoir une compréhension passive du français.

Le colloque s'organisera suivant trois axes thématiques qui ne correspondent pas forcément à des distinctions disciplinaires :

1. **Normes techniques :** Comment la transparence se traduit-elle sur le plan technique ? Quelles sont les techniques pour rendre le dispositif transparent à l'utilisateur ? A quoi correspond-t-elle du point de vue des pratiques des ingénieurs ? Quelles sont les aspects auxquels l'audit d'un système algorithmique devrait faire particulièrement attention ?

2. **Cadres épistémologiques :** Comment déterminer si ou quand un système technique est effectivement connu ? De quels critères disposons-nous pour rendre les systèmes lisibles ? Quels publics et quelles compétences sont présupposés par cette transparence ? Quelles sont les limites contextuelles ou structurelles qui empêchent d'accéder pleinement au fonctionnement du système ?
3. **Perspectives politiques :** Quelles exigences normatives devons-nous avoir envers la conception et la régulation des systèmes algorithmiques ? Comment ces exigences doivent-elles être informées par les normes techniques et les cadres épistémologiques ? Ces systèmes techniques ont-ils des spécificités qui réclament des modes de délibération et de régulation distincts ?

Références bilbiographiques indicatives

Wheeler, M. "The reappearing tool: transparency, smart technology, and the extended mind." *AI & Soc* 34**,** 857–866.

Ihde, Don, *Embodied Technics,* Automatic Press, United States, 2010.

Heidegger, Martin, *Etre et temps,* Editions Gallimard, Paris, 1986 [1927].

Grosman, Jérémy, & Reigeluth, Tyler, "Perspectives on algorithmic normativities: engineers, objects, activities." *Big Data & Society*, 2019.

Ananny, Mike, & Crawford, Kate,"Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability." *New Media & Society*, *20*(3), 2018.

Berns Thomas, *Gouverner sans gouverner. Une archéologie politique de la statistique.* Presses Universitaires de France, « Travaux pratiques », 2009.

Cardon Dominique, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, 2013/1 (n° 177), p. 63-95.

Pasquale, Frank, *The Black Box Society,* Havard University Press, Cambridge, MA, 2015.

Diakopoulos, Nicholas, « Algorithmic Accountability Reporting : On the Investigation of Black Boxes », Tow Center for Digital Journalism*, Columbia School of Journalism*, 2013.

Burrell, Jenna, « How the machine 'thinks': Understanding opacity in machine learning algorithms » in *Big Data & Society,* vol.3 n°1, 2016.

Sandvig, Christian, « When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software », *International Journal of Communication,* vol. 10, 2016, pp. 4972-4990.