Appel à propositions

« Éthique de l'IA » : enquêtes de terrain

Valérie Beaudouin (EHESS/CEMS) & Julia Velkovska (Orange Innovation/SENSE)

Les systèmes basés sur des algorithmes d'apprentissage - désignés souvent par la catégorie générique d'intelligence artificielle (IA), dont les contours restent cependant flous - sont présents dans un nombre croissant de sphères de la vie sociale. Ils affectent de plus en plus de pratiques sociales, comme les façons de s'informer, de se divertir, d'acheter, de nouer et maintenir des relations sociales, d'accéder à une formation, de rendre justice, de travailler, de recruter, de conduire, de réserver des vacances, d'accéder à des services et à des droits, etc. Ces technologies sont soutenues par les acteurs de l'innovation avec l'argument qu'elles seraient plus efficaces et plus neutres que l'être humain pour accomplir des tâches et pour rendre des services ou du moins qu'elles faciliteraient le travail et la vie quotidienne en offrant de l'aide.

La diffusion des outils algorithmiques, qui orientent les comportements, « prennent des décisions » ou entrent en interaction verbale avec les utilisateurs, s'est accompagnée d'une série de questionnements sur les risques associés. Les acteurs impliqués ont été les premiers à alerter sur les dangers (O'Neil, 2016, Dessalles, 2018). Ils ont été relayés par les médias qui ont pu amplifier et déformer les risques en jouant sur la peur du remplacement de l'homme par la machine (« laisseriez-vous un robot avocat vous défendre ? », « La menace grandissante des robots tueurs »…) mais aussi sur la défiance à l'égard des machines qui pourraient être injustes, biaisées (« *There's software used across the country to predict future criminals. And it's biased against blacks* »).

Peu à peu, les enjeux sociaux, économiques et éthiques de l'IA, ont été structurés par le milieu de l'IA autour d'un mot valise - « éthique de l'IA » (*AI ethics)* - qui désigne le pendant « social » de ce que « l'IA » représente aujourd'hui dans le domaine technologique, c'est-à-dire une catégorie générique pour désigner les technologies numériques. De la même façon la formule « éthique de l'IA » tend à englober l'ensemble des aspects « non-techniques » des systèmes algorithmiques, par exemple leurs usages, leurs conséquences sociales ou les régulations juridiques à envisager.

Sous le label « éthique de l'IA » sont assemblés toute une série de questions : celle des « biais » et du renforcement des stéréotypes (en raison des contenus sociaux et culturels incorporés dans les bases de données d'apprentissage), celle de l'opacité avec les problèmes d'interprétabilité et d'explicabilité des « décisions » prises par la machine, celle de la surveillance et de la protection des données personnelle, celle des risques de manipulations des attitudes et comportements (polarisation des opinions et propagation des discours haineux), pour ne citer que les principales. Ces questions ont émergé progressivement en mobilisant également les acteurs impliqués dans le développement même des technologies de l'IA.
En effet , de nombreux rapports traitant de « l'éthique de l'IA » ont été produits par des *think tanks*, des groupes d'experts ou des grandes entreprises du numérique qui visent à « éclairer » ou à orienter le législateur sur les politiques publiques et l'encadrement éthique des algorithmes. Cette profusion a largement contribué à la montée en puissance des discours sur l'encadrement éthique de l'IA dans l'espace public et à la mise à l'agenda médiatique et politique de ce sujet. Par exemple, les trois sujets prioritaires définis au lancement du Comité national pilote d'éthique du numérique (CNPEN) créé en France en 2019 étaient les agents conversationnels, le véhicule autonome et le diagnostic médical à l'ère de l'intelligence artificielle. Des initiatives similaires existent à l'étranger et au niveau des organisations

internationales comme l'UE et l'OCDE[1]. Ces innombrables rapports (plus d'une centaine en 2020) publiés autour de l'éthique de l'IA ont déjà fait l'objet de méta-analyses (Jobin *et al*., 2019 ; Schiff *et al*. 2020) et d'analyses critiques (Mittelstadt, 2019 ; Ganascia, 2017) et ont contribué à la constitution d'un sous-champ de recherche au sein de l'IA : AI Ethics.

Face à ce foisonnement de discours philosophiques, médiatiques, politico-administratifs et commerciaux sur « l'éthique de l'IA », les études empiriques qui examinent leur production, leur circulation, les formes d'appropriation et de performativité dans l'espace social sont rares. L'objectif de ce dossier de la revue *Réseaux* est de rassembler des articles qui proposent des entrées empiriques sur cette thématique, appuyés sur des enquêtes de terrain. De quoi « l'éthique de l'IA » est-elle le nom ? Comment prendre la mesure de la pluralité des discours et de l'hétérogénéité des pratiques se réclamant de cette catégorie (pratiques de lobbying, de marketing, de gestion des projets de conception, etc.) ?

Un premier volet du dossier adressera la question des discours et des controverses autour de l'éthique de l'IA dans l'espace public. Quelle est la généalogie de ces discours, comment s'organisent-ils, autour de quels arguments, portés par quels types de protagonistes ? Quels rôles jouent-ils dans les arènes médiatique, politique, scientifique et à leurs interfaces ? Forment-ils des controverses ? « L'éthique de l'IA » est-elle susceptible d'être constituée dans certains cas en problème public ?

Un deuxième volet réunira des études autour de dispositifs mobilisant de l'IA, qui soulèvent des questions « éthiques », depuis la conception jusqu'aux usages, dans une variété de domaines comme la médecine, la police et la justice (cf. par exemple le cas Compas aux États-Unis), la sécurité (par exemple la reconnaissance faciale), les réseaux sociaux, les agents conversationnels, etc. Comment les enjeux sociaux et éthiques sont-ils pris en compte par les concepteurs ? Par quelles opérations de traduction (d'un principe moral à un indicateur statistique) sont-ils intégrés dans la conception ? Comment les usages sont-ils évalués au regard de ces critères ?

Références

JOBIN Anna, IENCA Marcello et VAYENA Effy (2019), « The global landscape of AI ethics guidelines », in *Nature Machine Intelligence*, n° 9, vol. 1, p. 389-399.

SCHIFF Daniel, BIDDLE Justin, BORENSTEIN Jason et LAAS Kelly (2020), « What's Next for AI Ethics, Policy, and Governance? A Global Overview », *in AI Conference, February*, p. 153-158.

MITTELSTADT Brent (2019), « Principles alone cannot guarantee ethical AI », in *Nature Machine Intelligence*, n° 11, vol. 1, p. 501-507.

GANASCIA Jean-Gabriel (2017), *Le mythe de la Singularité. Faut-il craindre l'intelligence artificielle ?*, Paris, Seuil.

DESSALLES Jean-louis (2018), *Des intelligences TRÈS artificielles*, Paris, Odile Jacob.

O'NEIL Cathy (2016), *Weapons of Mass Destruction. How Big Data Increases Inequality and Threatens Democracy*, s.l., Crown, 259 p.

[1] AI HLEG (AI High Level Expert Group) (2018); *Ethic guidelines for trustworthy Artificial Intelligence*. Report for the European Commission. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai ; OECD Report (2019*). Artificial Intelligence in Society and AI Principles*

Calendrier prévisionnel

Nous vous demandons d'informer le secrétariat de rédaction de la revue de votre **intention de contribution** en adressant pour **le 1er septembre 2022,** une proposition d'article d'une ou deux pages précisant les questions de recherche, le corpus étudié et la ou les méthodes utilisées.

Les V1 seront à remettre le 2 janvier 2023.

Les intentions de contribution sont à adresser à : aurelie.bur@enpc.fr
Vous trouverez plus d'informations, notamment les consignes aux auteurs sur le site de la revue : http://www.revue-reseaux.fr/

La publication du dossier est prévue en septembre 2023.