

Compte-rendu de la 3^e Journée Éthique & Intelligence Artificielle

Grégory Bonnet – GREYC, Université de Caen Normandie

La 3^e Journée Éthique et Intelligence Artificielle, co-organisée par l'AFIA et le COERLE, s'est tenue le 2 juillet avec une quarantaine de participant·e·s (une trentaine le matin et une vingtaine l'après-midi), deux conférenciers invités et neuf orateur·rice·s.

1 Que signifie Éthique et IA ?

Une première question qui a été soulevée est celle du périmètre de ce domaine naissant, sachant qu'il convient peut-être plutôt de parler de systèmes intelligents que d'intelligence artificielle tant la sémantique de ce terme prend aujourd'hui une forme parfois étrange [1]. Traiter de l'éthique et des systèmes intelligents, est-ce s'intéresser aux questions éthiques ou est-ce faire de la recherche en éthique [3] ? Ou peut-être les deux, tout en prenant garde à « l'usage et l'abus du terme d'éthique, sous lequel nos contemporains entendent invariablement l'appel tous azimuts à la règle ? »¹ [7]. Quelle que soit la réponse, la diversité des exposés et des manières d'aborder les questions au cours de cette journée se sont articulées autour de trois thèmes :

- l'éthique des usages des applications
- l'éthique dans la prise de décision (roboethics)
- l'éthique dans la conduite de la recherche (responsible research)

2 Rapports, chartes et recommandations

En éthique, il n'y a jamais de réponses toutes faites mais plusieurs rapports et chartes présentés par les orateur·rice·s [1, 2, 3] peuvent servir de point d'entrée.

L'IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems [1] implique environ 1000 personnes mais peu de chercheur·e·s et encore moins de français·es. Son objectif est de faire en sorte que les concepteur·rice·s et programmeur·euse·s à tous les niveaux soient informés des questions d'éthique, et propose par le développement d'un programme universitaire sur la question éthique. L'Initiative a aussi publié un rapport² – Ethically Aligned Design – dont la version finale est prévue pour décembre 2018. Ce rapport se présente sous formes de fiches structurées : problématique, contexte, recommandations et références pour creuser les questions. Le processus de rédaction est fondé sur le consensus en tenant compte des cultures et préoccupations de chacun, sans pour autant tomber dans le relativisme éthique. Ces fiches serviront aussi peut-être de germes de standards qui développés par les comités de standardisation indépendants de l'IEEE.

Au niveau francophone, deux rapports [2] ont été produits ainsi que la charte³ Éthique et Big Data

1. Guillaume BADDOUAL. « L'éthique, elle aussi, est sans fond » (*Méditation d'une remarque d'Olga Alexandrovna Sedakóva*). Philosophie #116, pages 78–93, 2013. <https://www.cairn.info/revue-philosophie-2013-1-page-78.htm>

2. <https://ethicsinaction.ieee.org/>

3. <http://wiki.ethique-big-data.org>

[3]. Le rapport ⁴ de la CNIL, qui a hérité d’une réflexion sur l’éthique du numérique dans la cadre de la loi Lemaître, s’intitule Comment permettre à l’Homme de garder la main ? – et met en avant plusieurs principes fondateurs (loyauté et vigilance) et d’ingénierie (intelligibilité, responsabilité et intervention humaine). Une idée maîtresse est la mise en place d’une plateforme d’audit des algorithmes. Le second rapport ⁵ a été produit par la CERNA sur l’éthique de la recherche en apprentissage machine. Ce rapport identifie un ensemble de questions structuré en plusieurs catégories : quelles sont les données, les tâches de la machine, l’explicabilité, les décisions qui lui sont déléguées, les informations données aux utilisateurs et la responsabilité des concepteur·rice·s et programmeur·euse·s.

3 Des applications et des données

L’éthique est contextuelle et se pense en fonction des domaines d’application, et surtout sur le long terme. Plusieurs orateur·rice·s nous ont rappelé, qu’avec l’utilisation de masses de données, la frontière entre producteur·rice·s et utilisateur·rice·s devient floue [3, 5] et il qu’il convient de se doter d’une grille de lecture pour couvrir toutes les problématiques éthiques liées à une application.

Par exemple, dans le domaine du traitement automatique des langues, il ne faut pas oublier que les traductions des traducteur·rice·s humain·e·s sont souvent utilisées sans les rémunérer pour entraîner des machines et que nombreux sont les chercheur·euse·s à employer via l’Amazon Mechanical Turk des cohortes d’annotateur·rice·s humain·e·s payé·e·s au lance-pierre [3]. Dans le domaine médical, il ne faut pas oublier que la volonté de lutter contre les déserts médicaux peut conduire à l’implémentation de systèmes semblables aux cabines de diagnostic d’Axa qui ne garantissent pas aux patient·e·s un traitement digne et de confiance [5].

4 Décision et explications

S’il semble faire consensus qu’une machine – en soi – ne peut pas prendre de décisions éthiques mais les simule [1, 7], la question de la prise de décision selon des critères éthique et de l’explicabilité de ces décisions a été abordée dans le cadre des réseaux de neurones [4, 6].

S’il semble être une piste intéressante de coupler l’IA symbolique avec l’IA statistique [4], nos choix de représentation ont toujours une influence déterminante sur l’appréhension d’une situation préalable à toute décision éthique. En ce sens, une approche symbolique dans laquelle les catégories de perception sont pré-fixées est-elle un bon choix ? Inspiré du fonctionnement du cerveau, il est proposé de concevoir des couplages de réseaux de neurones afin de tenir compte des erreurs de bruit et d’être plus robuste dans l’identification d’une situation [6].

Au-delà des modèles, il serait aussi intéressant de travailler sur leurs limites (liées aux questions d’indécidabilité par exemple) et celles des systèmes décentralisés (que faire lorsque des agents ayant des hiérarchies de valeurs différentes interagissent ?) [7].

5 Table ronde « Jeunes chercheur·euse·s »

Pour conclure, une table ronde a été organisée entre trois jeunes chercheur·euse·s [8, 9, 10] qui travaillent sur des thématiques liées à l’éthique et l’intelligence artificielle. Trois perspectives de recherche ressortent de la discussion :

- Se servir de l’IA pour sensibiliser les utilisateur·rice·s à l’éthique,
- Travailler sur la modélisation de l’intention éthique,
- Penser les systèmes en termes de confiance car les utilisateur·rice·s sont vulnérables face à eux.

4. <https://www.cnil.fr/fr/mediatheque/rapports-annuels>

5. <http://cerna-ethics-allistene.org/Publications+CERNA/apprentissage/>

Références

- [1] Raja CHATILA.
Éclairage sur le rapport IEEE Ethically Aligned Design.
- [2] Claude KIRCHNER.
Éclairage sur les rapports de la CNIL et de la CERNA.
<https://ent.normandie-univ.fr/filex/get?k=wuhHKODjqq5gD0IX9jm>
- [3] Maxime AMBLARD et Karèn FORT.
Éthique et traitement automatique des langues.
<http://www.schplaf.org/kf/pdf/EthiqueEtIA.pdf>
- [4] Christophe DENIS.
Interprétabilité et validation d'applications métiers basées sur de l'IA statistique.
- [5] Joel COLLOC et Relwendé Aristide YAMEOGO.
Une boîte à outil éthique pour évaluer le statut des utilisateurs humains dans la relation de soin.
<https://ent.normandie-univ.fr/filex/get?k=6Mr6y2mJGLrXQIhzZq4>
- [6] Alexandre PITTI.
Contributions de l'IA incorporée et de la robotique développementale aux questions d'éthique.
- [7] Nazim FATES.
L'éthique peut-elle être mise en algorithme ?
<https://ent.normandie-univ.fr/filex/get?k=Q90c5L9BNheU5Bwac7a>
- [8] Azzedine BENABBOU.
Génération dynamique de situations critiques en environnements virtuels : dilemme et ambiguïté.
- [9] Fiona BERREBY.
Modélisation des systèmes éthiques.
- [10] Relwendé Aristide YAMEOGO.
Risques et perspectives des big data en santé et des systèmes d'information autonomes.