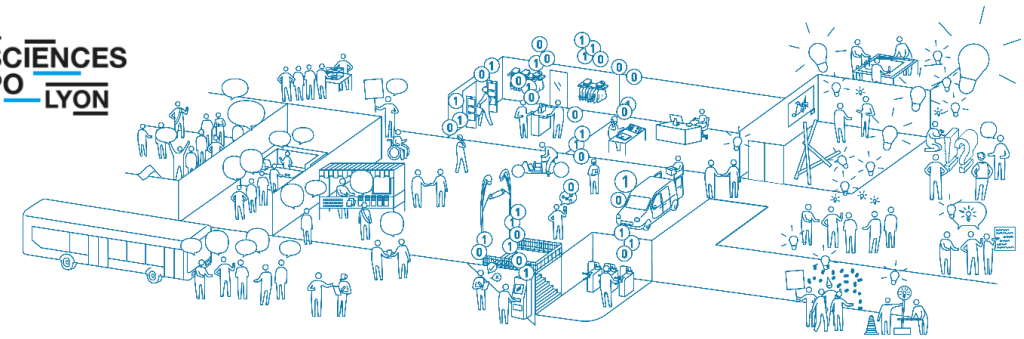

CHAIRE TRANSFORMATIONS DE L'ACTION PUBLIQUE

SÉMINAIRE « ALGORITHMES, INTELLIGENCE ARTIFICIELLE ET MONDE PUBLIC »

Séance 2 – Prospective des grandes fonctions de l'intelligence artificielle (IA) et de ses usages (18 novembre 2020)

Exposants invités : Dominique Cardon, sociologue et directeur du Medialab de Sciences Po. Renaud Vedel, coordonnateur de la stratégie pour l'intelligence artificielle.

Points clefs de la présentation de Renaud Vedel

L'histoire récente de l'IA : des évolutions majeures

L'IA est un programme scientifique et philosophique majeur dont l'un des objectifs principaux est de répliquer des fonctions cognitives humaines ainsi que de résoudre des problèmes habituellement traités par des humains. Bien qu'ayant déjà émergé aux débuts de l'informatique, ce programme a connu ces dix dernières années des évolutions notables. En effet, les années 2010 ont constitué un tournant au cours duquel sont survenues d'importantes avancées techniques qui peuvent être classées selon **3 dimensions** :

- **La perception :** qui concerne le traitement des signaux. En particulier, les images peuvent désormais être traitées et interprétées d'une manière beaucoup plus efficace qu'auparavant ;
- **Le traitement du langage naturel :** qui été le théâtre d'importantes améliorations, depuis 2017 notamment. Par rapport au traitement de l'image, ici, nous ne sommes plus uniquement du côté de la perception pure, mais également du côté de la cognition, de l'abstraction et – à un niveau certes encore basique - du raisonnement logique ;
- **La capacité de l'IA à intervenir dans son environnement :** une dimension associée à ce que Renaud Vedel appelle la « robotique environnementale ».

En combinant ces trois dimensions, on change drastiquement les potentialités d'un système d'IA ainsi que la nature des relations qu'entretiennent les humains avec celui-ci.

Les différentes applications fonctionnelles de l'IA

Plutôt que de traiter l'IA comme un ensemble uniforme comme c'est fait la plupart du temps, il est plus convenable de distinguer différents types **d'applications fonctionnelles**. Sous cet-angle là, il n'y aurait pas de raisons de séparer les usages de l'IA selon qu'ils se rapportent aux sphères publique ou privée, bien qu'il s'agisse de valeurs et de modes

économiques distincts. Cognilytica[1] a ainsi fait ressortir **7 applications fonctionnelles** (*patterns*) principales de l'IA :

- **L'hyperpersonnalisation :** la génération d'un profil individuel notamment utilisé à des fins de recommandation et de ciblage ;
- **La reconnaissance :** l'utilisation de l'IA pour structurer de l'information (images, texte, audio, etc.) et pour y identifier des éléments particuliers ;
- **La conversation et l'interaction avec l'humain :** permettre aux machines d'interagir avec des humains de la même manière que ceux-ci interagissent entre eux, que cela soit à travers des formes écrites ou parlées ;
- **Analyses prédictives :** qui ont pour objectif principal d'aider les humains à prendre des décisions ;
- **La détection de constantes et d'anomalies :** où il s'agit, pour des ensembles de données, d'identifier des schémas constants ou de cibler des éléments anormaux ;
- **Les systèmes autonomes :** la conception de systèmes capables d'effectuer des tâches ou d'interagir avec leur environnement tout en minimisant l'intervention humaine ;
- **Les systèmes tournés vers des objectifs précis (*goal-driven*) :** dont l'objectif principal est de trouver la solution optimale à un problème donné (*e.g.* simulation de scénarios, etc.).

Intelligence artificielle et confiance

Il est important d'intégrer des principes – entre autres éthiques - **dès la conception** d'un système d'IA. Ces principes doivent engager non seulement les ingénieurs, mais aussi plus largement l'ensemble des parties prenantes, dans le but ultime de consolider la confiance envers l'IA. Parmi les principes fondamentaux déterminés par le HLEG, on retrouve la **supervision humaine,** la **robustesse et** la **sécurité techniques,** la **confidentialité** et la **gouvernance des données,** la **transparence,** la **non-discrimination,** la **prise en compte des questions environnementales** et enfin la **responsabilité.**

Points clefs de la présentation de Dominique Cardon

Deux paradigmes pour deux visions de l'IA

À partir des années 1960, en informatique, deux grandes écoles paradigmatiques ont émergé aux États-Unis. Celle incarnée par l'ingénieur Douglas Engelbart avait pour idée directrice d'augmenter l'intelligence des humains par l'intermédiaire des machines. De l'autre côté, celle représentée par l'un des pionniers de l'IA, John McCarthy, avait pour ambition de produire des robots intelligents et surtout autonomes. C'est ainsi que, dès les années 1960, ces deux programmes ont porté deux idéologies distinctes qui persistent encore aujourd'hui : **rendre la société plus intelligente grâce à l'automate** pour le premier, et **mettre l'intelligence dans l'automate** pour le second. Toutefois, pour Dominique Cardon, c'est l'option de Douglas Engelbart qui finit toujours par l'emporter, étant davantage compatible avec l'idée selon laquelle on ne peut pas appréhender la technique de façon isolée. Car, comme l'illustre le développement de l'IA, **la technique, la société ainsi que nos représentations restent indissociables**.

Pour une définition plus restrictive de l'IA

[1] Outre ce travail de Cognilytica, d'autres grandes sources synthétiques de définition et classification des systèmes d'IA nous sont données par le rapport Principled Artificial Intelligence du Berkman Klein Center ainsi que par les travaux du High-Level Expert Group on Artificial Intelligence (références en fin de document).

L'usage courant de l'expression « intelligence artificielle » peut s'avérer problématique, puisqu'il a tendance à englober toutes sortes de systèmes informatiques souvent très différents. Les conséquences d'une telle définition extensive de l'IA s'observent notamment dans le champ éthique. En partant d'une conception large de l'IA on arrive de fait à une éthique trop générale, avec des règles et des principes qui sont peu applicables et peu concrets.

Pour être plus précis, il faut dire qu'il y a **non pas *une* mais plutôt *des* IA**, une seule d'entre elles étant réellement efficace aujourd'hui. Depuis les années 2000, **les réseaux de neurones** ont ainsi effectué leur retour sur le devant de la scène, alors qu'ils étaient pourtant ostracisés jusqu'il y a peu. C'est ce qui amène Dominique Cardon à privilégier une définition davantage restrictive de l'IA : « *un traitement de données avec des méthodes d'apprentissage plus ou moins sophistiquées, dont l'une d'entre elles, à savoir les réseaux de neurones, produit des résultats très étonnants* ». Par ailleurs, il est important de souligner que ces résultats appartiennent en grande majorité au registre de la perception et n'apportent pas pour le moment de nouveautés fondamentales du côté des capacités de raisonnement.

Catégoriser et calculer l'humain : les fonctions de l'État et les promesses de l'IA

L'une des grandes fonctions historiques de l'État est de **représenter la société par le haut avec des catégories**. L'humain est alors défini en étant **rangé** et **classé** dans des catégories, et des **corrélations** sont établies entre ces catégories, à l'instar de ce qui se fait en économie, en criminologie ou en marketing. Le fonctionnement de l'IA ainsi que celui des algorithmes reposent sur l'exploitation de très larges bases de données, des données qui sont issues du monde qui nous entoure. Sous cette perspective, l'IA et plus généralement les algorithmes peuvent être considérés comme **des systèmes techniques qui calculent le monde.** De son côté, le projet de ces systèmes techniques s'appuie sur une critique **individualiste** de la représentation catégorielle de la société dessinée par l'État. Au lieu de simplement mettre les individus dans une boîte, il s'agit de les calculer dans le cadre de leur singularité, en intégrant leur trajectoire de vie ainsi que leurs aspirations propres. La promesse de l'IA et des algorithmes est ainsi celle **de calculer les individus en dessous de ces catégories, de façon plus singulière et personnelle**. Un déplacement massif s'opère : plutôt que d'être calculés dans de grandes catégories comme celles que l'État a historiquement utilisées, on fabrique une société dans laquelle on personnalise suffisamment pour essayer d'être « mieux » calculés.

Il y a donc des **spécificités de l'État** à l'égard des calculs, des algorithmes et de l'IA. En analysant le fonctionnement de quatre exemples d'algorithmes utilisés par l'État, Dominique Cardon a identifié et classé ces spécificités selon **4 dimensions**. Souvent, l'État va mettre en place un algorithme pour **(1) résoudre des problèmes de justice.** On observe également dans ces utilisations étatiques **(2) un effet de centralisation.** Cela illustre une dynamique contradictoire des politiques publiques : d'un côté, on dit qu'il faut donner de l'autonomie à la décision locale, mais en même temps le *big data* et les scores prédictifs ont un effet de centralisation qui est important et qui se révèle problématique. D'autre part, ces dispositifs techniques se mettent souvent en place dans l'État en étant **(3) accompagnés d'une théorie sur la réduction des coûts,** et en intégrant des logiques de *néomanagement*. Enfin, la quatrième dimension est liée à l'attribution de ressources rares (allouer des places à l'université ou en crèches, des greffons, etc.). En ce qui concerne ces situations, l'une des manières pour l'État de justifier ses politiques publiques a été de définir clairement les critères d'attribution. Mais cette définition vacille avec l'IA et les algorithmes, qui font entrer **(4) l'évaluation du futur à l'intérieur même du calcul et de cette définition**. À présent, l'estimation de la réussite des politiques publiques pour tel ou tel type d'individu intervient pour définir les critères initiaux qui sont offerts au calcul.

Pour aller plus loin …

Cardon, D. (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*. Seuil.
Cognilityca. (2019). *The 7 Patterns of AI*. https://www.cognilytica.com/2019/04/04/the-seven-patterns-of-ai/
Commission Européenne. (2018). *High-Level Expert Group on Artificial Intelligence*. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence
Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence : Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (SSRN Scholarly Paper ID 3518482). Social Science Research Network. https://papers.ssrn.com/abstract=3518482
HLEG. (2019). *Lignes directrices en matière d'éthique pour une IA digne de confiance*. http://op.europa.eu/fr/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/prodSystem-cellar/language-fr/format-PDF
OCDE. (2019). *Artificial intelligence*. https://www.oecd.org/going-digital/ai/