



CHAIRE TRANSFORMATIONS DE L'ACTION PUBLIQUE

SÉMINAIRE « ALGORITHMES, INTELLIGENCE ARTIFICIELLE ET MONDE PUBLIC »

Séance 1 - État de l'art, état des forces. Qu'est-ce que l'intelligence artificielle en 2020 ? (7 octobre 2020)

Invités : Daniel Le Métayer, directeur de recherche à INRIA. Jacques-François Marchandise, co-fondateur et délégué général de la Fondation Internet Nouvelle Génération (Fing).

Points clefs de la présentation de Daniel Le Métayer

Mise au point technique et sémantique : algorithmes, apprentissage et intelligence artificielle (IA)

Le discours médiatique et public autour des questions relatives à l'IA et aux algorithmes se caractérise par une forte polysémie, ces expressions ayant tendance à être utilisées de manière interchangeable. Néanmoins, les termes d'IA et d'algorithme ne renvoient pas tout à fait au même objet.

Pour le dire simplement, la notion d'**algorithme** correspond à une suite d'étapes et de règles plus ou moins clairement établies et appliquées en vue de la résolution d'un problème. Sur le plan opérationnel se distinguent toutefois deux grandes familles au sein de la programmation informatique.

Un premier paradigme dit « standard » consiste à inscrire la logique de l'algorithme directement dans le programme informatique.

Le second paradigme est quant à lui centré sur la notion d'**apprentissage**. Ici, il s'agit d'alimenter le système avec des données permettant la génération d'un modèle informatique qui servira à structurer ainsi qu'à agencer les informations reçues.

De son côté, l'**IA** désigne - de manière générale - l'utilisation de techniques informatiques pour réaliser des tâches qui relèvent habituellement de l'intelligence humaine. Bien qu'elle y soit souvent assimilée, l'IA ne correspond pas tout à fait à l'apprentissage, puisqu'il est par exemple possible de parler d'IA « symbolique », où les règles sont explicitement énoncées.

Quels éléments d'analyse pour les systèmes d'apprentissage et sous quelles dimensions peut-on caractériser ces derniers ?

Trois éléments - techniques, mais pas que - majeurs sont à prendre en considération lorsque l'on parle d'apprentissage. Les **données en entrée** représentent le point de départ de tout apprentissage informatique et conditionnent le fonctionnement effectif de celui-ci. Le second élément concerne le **procédé d'apprentissage utilisé**, qui se décline en plusieurs catégories génériques (supervisé, non-supervisé, continu, etc.).

Enfin, la **représentation du résultat** est une composante importante où il est possible de choisir différentes formes et méthodes, qu'il s'agisse d'arbres de décision ou de réseaux de neurones organisés en couches.

Les systèmes d'apprentissage qui viennent d'être évoqués peuvent être caractérisés selon plusieurs dimensions et critères. La **précision** d'un système peut par exemple s'exprimer en fonction de la proportion de faux positifs et de faux négatifs parmi les résultats. Complémentaire de la précision, la **robustesse** du système nous renseigne sur sa capacité à être induit en erreur plus ou moins facilement, par exemple lorsque les données en entrée sont légèrement modifiées. L'**absence ou la présence de biais**, notamment discriminatoires, est aussi un critère fréquemment utilisé pour caractériser un système et évaluer la représentativité de ses données d'apprentissage. L'**intelligibilité des résultats** représente quant à elle une dimension primordiale permettant de juger de la transparence ou au contraire de l'opacité d'un système d'apprentissage.

Algorithmes d'aide à la décision dans le monde public, une diversité d'enjeux

En plus d'occuper une place de plus en plus importante dans notre quotidien, les algorithmes d'aide à la décision se sont largement immiscés dans le paysage institutionnel. Pour ne citer quelques exemples : sélection à la sortie du lycée, mutation du travail des enseignants, accès à l'emploi, ou même encore dans le domaine de la justice. Cette omniprésence n'est pas sans soulever de grands enjeux tels que celui de **la protection des données personnelles et de la vie privée**, celui de la **fiabilité** et de l'**efficacité**, celui de la **discrimination** et des **biais**, celui de l'**intelligibilité** et de l'**explicabilité des résultats** ou encore celui de la **légitimité** même des algorithmes d'aide à la décision et de leurs applications.

Points clefs de la présentation de Jacques-François Marchandise

Quelle intelligence artificielle voulons-nous ?

Nous avons assisté, depuis plus d'une vingtaine d'années maintenant, au passage d'une période relativement pionnière des nouvelles technologies à celle de leur massification et de l'augmentation de leurs effets de socialisation.

Aujourd'hui, l'intelligence artificielle recouvre une partie en croissance permanente de nos mondes sociaux, que l'on parle d'environnements politiques, de rapports de force ou de façons de sociabiliser. Plusieurs questions aussi légitimes que centrales surgissent alors : est-ce que le numérique du début s'est éloigné de ses promesses initiales ? Et de manière plus fondamentale : quel numérique et de quelle IA voulons-nous ?

Dans la perspective d'apporter des éléments de réponse et de scruter les systèmes et dispositifs sociotechniques actuels, un prisme d'analyse **tridimensionnel** peut être utilisé. Ce prisme combine trois angles, à savoir celui de l'**offre** (les parties qui font et conçoivent le numérique), celui de la **commande** (les parties qui en sont les commanditaires et qui portent les financements, les stratégies publiques ou encore les investissements), et enfin celui des **usages** (ce que les utilisateurs et plus généralement les citoyens ont en face).

Lorsque l'IA déraile

En matière d'IA et d'usages publics, il est frappant de constater que la majorité des discussions spontanées autour de l'IA recourent de façon quasi-systématique à des exemples où celle-ci pose problème. On a tous en tête un cas de figure où l'IA ne marche pas bien et déraile. Par exemple, lorsqu'une des GAFAM met un an pour supprimer des biais sexistes et racistes de son système de recrutement automatisé. Ou, sur le terrain des politiques

sociales, lorsque l'IA accentue la surveillance des personnes les plus précaires alors qu'elle était censée améliorer leur accompagnement.

Comment rendre compte de cette situation ? Plutôt que d'opter pour l'hypothèse de l'accident, l'analyse des modèles économiques ainsi que des modes de socialisation et de diffusion propres au numérique révèle une succession permanente et frénétique de « prototypes ». Dans un système d'innovation permanente, les dispositifs numériques relèvent souvent d'un fonctionnement « quick-and-dirty », où l'on privilégie la rapidité d'une solution pour résoudre un problème donné. Ce n'est généralement que lorsque ces dispositifs numériques ont un impact direct et explicite sur la vie humaine que l'on se soucie davantage de leurs enjeux politiques, sociaux et éthiques.

Technique, pouvoirs et contre-pouvoirs

Quelles sont aujourd'hui, en termes d'IA, les commandes les plus fréquentes des acteurs publics ? Ces commandes concernent notamment la sécurité, la surveillance ou encore le contrôle (production de tableaux de bord, de prédiction, etc.). La puissance en augmentation permanente de ces dispositifs ne rend que plus importants les problèmes qui vont se poser. La technique c'est du pouvoir. Les techniques numériques ne dérogent pas à cette règle : elles représentent du pouvoir technologique, du pouvoir économique, et du pouvoir politique. L'IA, comme le pose Kate Crawford, est une nouvelle ingénierie du pouvoir. Mais où sont alors les contre-pouvoirs potentiels de ces systèmes fondés sur le contrôle des individus ? Comment ne pas aller tout droit dans une société de contrôle et de surveillance (banalisation de la reconnaissance faciale, du fichage, etc.) ?

Acteurs territoriaux : la question de la maîtrise d'ouvrage et de la culture des données

Dans un ensemble de collectivités territoriales et parmi les acteurs en charge de politiques publiques, le sentiment d'un désarmement face à la puissance et à la technicité des outils technologiques qui sont proposés est souvent partagé. Ce sentiment de désarmement se traduit par un doute sur la propre capacité de ces acteurs à maîtriser les outils qu'ils sont en train de mettre en place (capteurs sur la voie publique ou au domicile des personnes, internet des objets, etc.). En ce sens, sur plusieurs usages séduisants ou pertinents de l'IA, le régime qui s'est installé est un régime du doute et de l'inquiétude. Des doutes et des inquiétudes émanant des exemples de l'IA qui « déraillent », mais aussi des doutes et des inquiétudes quant aux compétences requises pour une bonne maîtrise d'ouvrage et d'usage.

Des questions additionnelles émergent : est-ce qu'une maîtrise d'ouvrage publique de l'IA est possible et que faire quand elle ne l'est pas ? De la même façon, comment décrire les compétences nécessaires pour les non-informaticiens en face de l'IA ? L'enjeu ici n'est pas tellement l'ouverture des données, mais la culture (ou littératie) des données. Aujourd'hui se pose la question de comment avoir la culture des algorithmes, car nous aurons besoin de demander des comptes à des systèmes qui prennent des décisions sur nous sans nous.

Internet des objets et investissements publics : une absence de principes communs

Sur la question de l'internet des objets, on constate qu'aujourd'hui l'achat public manque de principes d'action communs lors des investissements sur l'internet des objets. Que cela soit par rapport à la question de l'impact environnemental, de la collecte des données, du respect de la vie privée, il n'y a pas assez de coordination et de concertation. Comme pour l'IA, la responsabilité se retrouve alors polarisée du côté des concepteurs. Le constat dressé est souvent le suivant : ce n'est que rarement que les commanditaires exigent le respect des principes qui viennent d'être évoqués, les concepteurs prenant toutes ces questions en charge - lorsque c'est le cas - de manière « quasi-héroïque ».

Pour aller plus loin ...

- CNIL. (2017). Comment permettre à l'Homme de garder la main? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf
- Eubanks, V. (2018). Automating Inequality : How High-tech Tools Profile, Police, and Punish the Poor. St Martin's Press.
- Guillaud, H. (2019). Kate Crawford: « l'IA est une nouvelle ingénierie du pouvoir ». internetactu. <http://www.internetactu.net/2019/09/30/kate-crawford-lia-est-une-nouvelle-ingenierie-du-pouvoir/>
- Le Métayer, D., & Desmoulin-Canselier, S. (2020). Décider avec les algorithmes : Quelle place pour l'Homme, quelle place pour le droit ? (1ère ed.). Dalloz.