fellows

le regard de chercheurs internationaux sur l'actualité

n°40

1er mai 2018

http://fellows.rfiea.fr

Réseau français des instituts d'études avancées
Aix-Marseille • Lyon • Nantes • Paris

Intelligence artificielle

La justice phagocytée par les algorithmes

Une intelligence artificielle « neutre » est-elle possible ?

Aleš Završnik

[Eurias, Collegium Helveticum, 2017-2018]

Aleš Završnik est chercheur à l'Institut de criminologie et professeur associé à la faculté de Droit de l'Université de Ljubljana (Slovénie). Il a été expert en éthique au Conseil européen de la Recherche. Ses recherches portent sur la criminalité et la technologie, le droit des technologies de l'information, la cybercriminalité, la surveillance et les méfaits sociaux des technologies. Il a notamment publié *Big data, crime and social control* (Routledge, Londres, 2018) et prépare actuellement un ouvrage sur les périls de la justice algorithmique.

*Programme européen coordonnée par la fondation RFIEA

Zavrsink

VERS UNE JUSTICE AUTOMATISÉE ?

Notre monde fonctionne à bien des égards grâce aux *big data* (données massives), aux algorithmes et à l'intelligence artificielle (IA) : les réseaux sociaux nous suggèrent avec qui devenir ami, les algorithmes placent nos actions en bourse, et notre vie amoureuse n'est plus une zone dénuée de statistiques. Le couple *big data*-algorithmes est devenu une problématique cruciale en matière de renseignement, de sécurité, de défense, d'anti-terrorisme et de politique criminelle : les ordinateurs aident l'armée à trouver ses cibles et, comme l'a révélé Edward Snowden en 2013, les agences de renseignement étayent leurs analyses sur la base d'une surveillance préventive de masse des réseaux publics de télécommunications. Les algorithmes qui permettent d'interpréter de façon intelligible les *big data* constituent un nouveau type de production de connaissances dans le domaine de la lutte contre la criminalité. Pour décider de l'affectation de leurs moyens, les forces de l'ordre utilisent de plus en plus de logiciels de prévision des crimes. De son côté, la justice pénale s'appuie de façon grandissante sur des algorithmes et des instruments de prédiction de la peine.

Alors que les informaticiens sont préoccupés par les exigences croissantes en matière de capacité de calcul, de stockage et de communication (elles atteignent les limites des lois physiques en termes de fiabilité et d'économie), les chercheurs en sciences sociales de leur côté n'ont pas encore pleinement saisi les implications sociales et éthiques des paradigmes informatiques existants. La question n'est pas seulement de savoir jusqu'où ira (et devrait aller) l'automatisation de la gouvernance grâce à ces nouvelles solutions informatiques très puissantes, telles que l'informatique quantique, bio-inspirée et participative ; elle est aussi de déterminer quelles sont les conséquences socialement destructrices des *big data* et des systèmes automatisés de prise de décision – qui sont déjà à l'œuvre dans les sociétés capitalistes contemporaines basées sur la surveillance. Nous pouvons déjà observer des discriminations qui touchent en premier lieu les couches les moins aisées et les moins puissantes de la population, par exemple dans les domaines de l'assurance, de l'éducation, de l'emploi, de la criminalité et de la sécurité, et on constate de nombreuses distorsions des processus

Les frontières entre les concepts de « suspect », d'« accusé » et de « condamné » commencent à disparaître.

démocratiques, par exemple sur la base de données de réseaux sociaux, comme l'a révélé les dénonciateurs de Cambridge Analytica en 2018.

Criminalité et sécurité

Les processus décisionnels automatisés empiètent fortement sur les libertés fondamentales dans le cadre pénal et policier, car ces acteurs détiennent le monopole de l'usage légal de la force physique. Au cours des siècles suivant les Lumières, des concepts et des procédures juridiques très nuancés ont été conçus pour réglementer le recours à la force. Ce qui est en jeu avec les processus décisionnels automatisés grâce à l'IA est le fondement même des concepts et procédures juridiques pensés pour réguler ce recours. Aujourd'hui, les forces de l'ordre n'opèrent plus seulement dans le paradigme du système de punition ex-post facto fondé sur des actes criminels manifestes ; elles utilisent de manière croissante des mesures préventives ex ante, basées sur des états psychologiques présumés, comme la notion de « terroriste dormant » dans la législation anti-terroriste allemande de l'après-11 septembre, qui est un exemple édifiant d'identification « algorithmique » d'une « cible » (par la suite invalidée par la Cour constitutionnelle fédérale).

Lorsqu'à des fins de profilage ou pour prédire le lieu et l'heure de futurs locus delicti commissi, les forces de l'ordre et les services de renseignement collectent des données - ou confient cette mission à des géants des télécom en échange d'un régime réglementaire complaisant - les frontières entre les concepts de « suspect », d'« accusé » et de « condamné » commencent à disparaître. Lorsque les logiciels de prévision policière utilisent des statistiques concernant tous les types de crimes, la petite délinquance attire davantage l'attention que les « crimes en col blanc ». Certes, les délits mineurs aident PredPol – logiciel de prédiction de délits – à prévoir les coordonnées GPS des infractions graves, mais à quel prix ? **La police se concentre de façon disproportionnée sur les minorités qui, par conséquent, sont « sur-policiarisées » : en portant une attention accrue à un groupe spécifique, la police y détecte davantage de crimes, ce qui génère des données incitant à amplifier encore le contrôle policier. Les crimes financiers ou autres « crimes en col blanc » sont, quant à eux, négligés de façon totalement disproportionnée par rapport à l'ampleur des dommages causés.** La justice pénale utilise également de plus en plus d'outils décisionnels d'IA pour pronostiquer les futurs crimes des personnes en attente d'un procès ou d'une libération conditionnelle, mais aussi pour prédire le refus de collaborer avec les autorités dans les procédures de mise en liberté sous caution. L'algorithme de la fondation Arnold, qui est utilisé dans 21 juridictions aux États-Unis, brasse ainsi 1,5 million d'affaires pénales pour prévoir le comportement des accusés. **Une étude menée par des chercheurs de l'Université Stanford à partir 1,36 million de cas de détention provisoire assure qu'un ordinateur peut prédire si un suspect fuira ou récidivera mieux qu'un juge humain. Mais données et algorithmes sont des artefacts humains. Les systèmes automatisés de prise de décision sont donc susceptibles d'amplifier considérablement les erreurs et les défauts humains.** Comme avec les « flash crashes » du trading à haute fréquence, où 99% de la valeur d'une action peut être réduite à néant en quelques minutes, les systèmes de décision automatisés peuvent envoyer de manière disproportionnée certains groupes en prison. Se fier excessivement aux calculs automatisés du risque peut déclencher un cercle vicieux de mauvaises décisions et exacerber les problèmes sociaux existants. Par exemple, dans une évaluation détaillée de l'algorithme de récidive COMPAS, ProPublica a découvert que le système était biaisé contre les personnes noires. Plusieurs chercheurs ont d'ailleurs mis en garde contre la façon dont la « gouvernance automatisée » ou la « gouvernementalité algorithmique » pouvait perpétuer et amplifier les discriminations. Cela a déjà empiété sur les libertés fondamentales, dont le droit à la non-discrimination fait partie, et reproduira inévitablement les inégalités, puisque les données recueillies sont divisées selon des critères économiques, raciaux, ethniques et de genre. Dans le cadre de la justice pénale, l'intelligence artificielle porte atteinte à l'égalité des parties dans les procédures judiciaires, ainsi qu'au droit à un procès équitable, c'est-à-dire le droit à un juge humain ainsi qu'à un tribunal indépendant et impartial. La procédure pénale doit concilier équité et efficacité, mais l'équité a toujours primé : mieux vaut laisser dix criminels s'évader de prison que de condamner un innocent. Cela a longtemps été l'une des principales distinctions entre les systèmes politiques autoritaires et démocratiques.

De la démocratie à l'« algocratie » ?

Une expérience menée sur la contagion émotionnelle massive des utilisateurs de Facebook a démontré comment des outils puissants induisent l'humeur et l'opinion. Le récent cas de Cambridge Analytica a montré que de puissants outils existent aussi pour induire une « contagion politique » auprès du grand public. La « justice automatisée » n'est qu'une part infime d'une tendance plus large vers une « gouvernance automatisée » susceptible de fausser les processus démocratiques. Ce qui est en jeu avec l'avènement d'outils de prise de décision basés sur l'IA, c'est l'État de droit, lentement remplacé par la « règle de l'algorithme ». Avec, en perspective, une algocratie supplantant insidieusement la démocratie.

Pour aller plus loin

Retrouvez l'article d'Aleš Završnik, ses références et des contenus complémentaires sur **fellows.rfiea.fr**

Ian Davidson

[Collegium de Lyon 2017-2018]

IS A "NEUTRAL" ARTIFICIAL INTELLIGENCE POSSIBLE?

Does the development of artificial intelligence (AI) necessarily mean greater control and infringement of personal data, fundamental freedoms and privacy? Or is it possible to move towards artificial intelligences that do not attack these dimensions or, even, protect them?

"Technology is neither good nor bad; nor is it neutral". With this now classic statement, the technology historian Melvin Kranzberg tells us that it is only what we, humans, do with the technology that determines its moral impact. The very same AI can have different results when introduced into different situations. Many current applications by AI companies have gained notorious press coverage such as the improper use of personal data (where 87 million Facebook users' data was shared with Cambridge Analytica), and censorship of search results (by internet search companies such as Google). But companies such as Facebook and Google are allowing you to use their products for free for ulterior reasons, not just to serve your interests. In particular, many internet AI companies wish to perform better targeted advertisements: the advertisement dollars they earn pay for their programmers, data farms and other aspects of their core business. This means **the previously mentioned transgressions and other infringement of personal data, freedoms and privacy seem simply inevitable.** Advertisers have many options and in turn organizations attempting to gain these dollars will push the ethical boundaries of privacy and related topics.

However, as Kranzberg mentions the same technology can produce different results in different contexts. Consider two very different contexts: i) Personal AI assistants and ii) AI prisoner risk assessments. AI assistants such as Amazon's Alexa, Google's Allo and Apple's Siri are so far one of the most pervasive applications of AI with the ability to impact billions of people on a daily basis. Today the focus of these tools is simple: understanding what music to play, adding to our shopping list or sending text messages. However, the long term vision of these assistants is to hold conversations with a user and be delegated tasks, which entail making decisions on our behalf. This may seem far-fetched, but naysayers can look at the self driving cars of today and compare the progress over just the last five years. Such assistants are an inevitable application of AI, as they effectively allow ourselves to be extended. But just as advertisers make demands to get better results, individuals will demand the need for the AI to be neutral. Clearly we all want the AI assistant to make decisions for the benefit of us, not for the advertisers. Hence, just as it is inevitable that advertisement based products will always be on the border of infringing our privacy, these types of personal assistants must be neutral if they are to exist. And the demand for them is so great they will exist.

Consider the *risk assessments* for those charged with crimes. Risk assessments involve determining the person's propensity to be a repeat offender or even propensity for violent crime. These risk assessments are used in many ways, including setting bonds and sentencing. The US Justice Department's National Institute of Corrections encourages this information to be used and it is currently used in Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington and Wisconsin. Such AI technologies should meet immense societal demands to be neutral, and will have to be under constant scrutiny.

What are the technical challenges to overcome?

Ian Davidson

Ian Davidson est professeur titulaire d'informatique à l'université de Californie - Davis. Son programme de recherche est consacré aux contributions algorithmiques fondamentales à l'intelligence artificielle, à l'apprentissage automatique et à la science des données. Il s'intéresse à leur application dans des domaines ayant des impacts sociétaux tels que les neurosciences, les systèmes de tutorat intelligent et les réseaux sociaux. Il privilégie pour cela la collaboration avec des chercheurs d'autres disciplines. Il a notamment obtenu le prestigieux prix Career de la National Scientific Foundation et est auteur d'une quinzaine d'articles scientifiques dans les plus grandes revues internationales.

Davidson2017©ChDelory

AI companies are addressing these challenges in three different ways: (i) explainable AI, (ii) fair machine learning and (iii) differential privacy. Explainable AI (XAI) offers the possibility of moving the AI from a black box to be explainable to a human. This involves the AI not just making decisions or predictions but also augmenting the decision/prediction with answers to auxiliary questions such as: ``Explain why you made this prediction, and why not this other prediction", "Explain when and why your are most/least confident" and "Explain when you are likely to make mistakes". Many AI systems also involve machine learning. This involves teaching a machine to solve a task such as predicting if a loan will be successfully paid back, or predicting who would be a good hire for a job. Machine learning works by being provided many positive and negative examples and then allowing the computer to learn a predictive model based on this data. But **if a machine is taught from examples that reflect the stereotypes found in human culture then it will be biased just as the data is biased. The need for unbiased machine learnt models has produced the emerging field of fair machine learning.** Finally, one of the greatest challenges of AI systems is related to domains with highly sensitive information such as medical records. Differential privacy aims to perform computations such as aggregate queries and computations without the ability to identify individuals from which the computation was calculated. Extensions which give the user even more privacy such as local differential privacy, have been embraced by Google to collect web browsing information (using their RAPPOR system), Apple to collect typing history and Microsoft to collect telemetry information over time.

If AI machines get to think more and more as humans do, shall we reconsider their legal status? Will they remain simply machines?

There has been considerable debates on the rights of AI machines. In 2017, the EU parliament passed a resolution to the European Commission on civil law rules that recommends sophisticated robots be given "electronic persons" legal status. The report states "so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to case where robots make autonomous decisions". The argument for such rights is the following: if AI robots are going to replace humans at complex tasks such as surgery, they should be held responsible, as a human would. However, the matter is not so straightforward. The major opposition to this granting of legal status to robots lies in that if a robot is given legal status based on the Natural Person model, then it must be given other rights such as dignity or privacy, directly confronting and competing with human rights. A group of several hundred AI, robotic and ethics experts have just signed a letter in opposition to this potential move.

Pour aller plus loin

Retrouvez l'article d'Ian Davidson ainsi que des contenus et références complémentaires sur **fellows.rfiea.fr**

4 instituts d'études avancées en réseau
IMéRA, IEA d'Aix-Marseille
Collegium de Lyon
IEA de Nantes
IEA de Paris

Direction éditoriale
Olivier Bouin
Philippe Rousselot

Contactez-nous !
Fondation RFIEA
Julien Ténédos
Aurélie Louchart
contact@rfiea.fr
01 40 48 65 57

rfiea.fr
54 bd Raspail
75006 Paris

En bref

COLLOQUE

Le sociologue **Daniel Mercure**, résident 2017-2018 de l'**IEA de Paris**, organise une journée de réflexion sur « Les transformations contemporaines du rapport au travail », en collaboration avec l'Université Paris-Saclay. L'objectif de ce colloque est de repérer, de contextualiser et d'analyser les principales lignes directrices des transformations du rapport au travail dans les sociétés occidentales au cours des trente dernières années. Assistons-nous à une réelle reconfiguration du rapport au travail ? Selon quelles modalités ? Sous l'impulsion de quels facteurs de contingence ? Quels effets sur la vie des travailleurs ?

Le 24 mai 2018
9h à 18h30
IEA de Paris Hôtel de Lauzun
17, quai d'Anjou
75004 Paris

PUBLICATION

Alain Supiot, professeur au Collège de France, fondateur et membre émérite de l'**IEA de Nantes** vient de publier *Face à l'irresponsabilité : la dynamique de la solidarité*. L'ouvrage s'intéresse à l'idée de « responsabilité solidaire », qui oblige ceux qui ont le pouvoir économique à répondre légalement des conséquences de leurs décisions. L'essai comprend des contributions de **Samuel Jubé** (directeur de l'IEA de Nantes), **Samantha Besson** (membre du conseil scientifique et du comité stratégique de l'IEA de Nantes) et **Jeseong Park** (résident 2011-2012 de l'IEA de Nantes).

Face à l'irresponsabilité : la dynamique de la solidarité, éditions du Collège de France, 2018, 184 p.

JOURNÉE D'ÉTUDE

L'**IMéRA d'Aix-Marseille** organise une journée d'études autour des projets lauréats de l'appel d'offre écosystèmes continentaux et risques environnementaux (ECCOREV) de 2016. **Raouf Boucekkine**, directeur de l'IMERA et **Anna Serra Llobet**, chercheuse à l'université de Californie - Berkeley et résidente IMéRA 2017-2018, ouvriront la journée avec une conférence intitulée : « Managing flood risk behind levees : what US can learn from France ». Les porteurs de chaque projet présenteront leurs résultats et débattront avec la salle.

Le 23 mai 2018
9h à 17h
IMéRA Maison des astronomes
2, place Le Verrier
Inscription obligatoire
sur le site d'ECCOREV
www.eccorev.fr