

Conférence du 6 avril 2017, 13h à 16h, Sciences Po Lille :

Le Choix à l'heure du Big Data – conférence suite à l'appel à la réflexion collective de la CNIL : « Ethique et Numérique, les algorithmes en débat »

1. François Coupez

(cofondateur d'ATIPIC, avocat, spécialisation en droit des nouvelles technologies, de l'informatique et de la communication, enseignant à Paris II)

L'algorithme a une véritable valeur marchande, puisqu'il permet de faire la différence avec d'autres entreprises.

Un algorithme est une suite d'instructions permettant de résoudre un problème ou d'obtenir un résultat. On peut le comparer à une recette de cuisine : si on ne suit pas les instructions, le résultat ne va pas correspondre aux attentes d'origine. Un algorithme peut ensuite être utilisé pour construire un logiciel – si l'algorithme est le squelette, le logiciel est comparable au corps.

La question de la protection juridique de l'algorithme :

- Quand on développe une nouvelle idée, on est protégé par le droit d'auteur et le droit de la propriété intellectuelle. On ne peut pas s'approprier les idées qui sont de libre parcours. On ne peut pas non plus s'approprier une méthode, et donc un algorithme. Par conséquent, l'algorithme n'est pas protégeable.

Mais il est possible de protéger un logiciel, dans lequel est intégré l'algorithme. En effet, la protection via le droit d'auteur est possible à partir du moment où il y a 1) une formalisation et 2) l'originalité.

En matière de logiciel, il est très difficile d'apporter la preuve de l'originalité. C'est pourquoi les tribunaux utilisent des principes, cf ccass, 16 mars 86 : fait d'avoir exercé ses choix à partir du moment où l'algorithme laissait la possibilité de faire plusieurs choix, et savoir expliquer ces choix.

- L'autre problématique est celui du droit des brevets qui nécessite la nouveauté, l'application industrielle et l'activité inventive. Le droit des brevets ne peut s'appliquer aux algorithmes que s'ils font partie d'un procédé industriel plus large.

Comment est protégé l'algorithme ? Par le secret via des accords de confidentialité très stricts → contraire à l'idée de brevet. => mais donc on ne sait pas comment l'algorithme fonctionne.

Pour comprendre les résultats que les algorithmes donnent, il est important de savoir comment il est nourri, quels critères sont pris en compte. Or, cela est secret, ce qui s'oppose à la demande de transparence de plus en plus fréquente. → On ne peut pas contrôler si on ne connaît pas le fonctionnement.

Pourtant ces algorithmes peuvent avoir un impact important dans la vie de tous les jours.

Exemple du robot (intelligence artificielle) de Microsoft qui peut mener des conversations : 24h après le lancement de l'outil sur twitter, un groupe d'activistes a perverti son fonctionnement, en l'abreuvant d'idées extrémistes. Or, l'algorithme était conçu pour être nourri par les informations qu'on lui donnait, indépendamment de la source. Il n'était pas conçu pour se protéger contre les abus.

La réaction du législateur :

Il y a une demande de transparence.

La loi pour une République numérique a été promulguée en octobre 2016, mais il y a des lois plus anciennes dans le code monétaire et financier (article L 451-4), qui datent de 2013. Elles concernent le Trading à haute fréquence → Les traders humains sont de plus en plus remplacés par des algorithmes pour passer des ordres quasiment à la vitesse de la lumière. Cependant, les algorithmes sont à l'origine des flash crash. En effet, alors que les humains agissent raisonnablement dans les moments tendus, les algorithmes renforcent l'effet moutonnier et rajoutent des tensions.

Réglementation en 2013 :

Objectif : transparence → cette réglementation prévoit que si on utilise des systèmes de disposition automatisée générant des ventes, il faut le notifier à l'AMF. On doit assurer une traçabilité de chaque ordre envoyé & conserver les algorithmes → de cette façon, des enquêtes peuvent être menées après coup si nécessaire.

Article 311-3-1 du code des relations entre le public et l'administration : A partir du moment où une décision individuelle est prise sur le fondement d'un traitement algorithmique, l'intéressé doit en être averti, sans qu'on doive aller dans le détail du fonctionnement de cet algorithme.

Le problème : moins l'on en sait sur le fonctionnement de l'algorithme, plus il est efficace, ce qui s'oppose à la demande de transparence => il faut donc trouver un compromis. Tout l'enjeu est d'augmenter la transparence même si cela implique de baisser le niveau d'efficacité des algorithmes.

Blockchain :

Elle est au fondement du Bitcoin (crypto monnaie). Ce qui détermine le fonctionnement d'ensemble c'est le code, le développement informatique que personne ne détient.

Problème : « Code is law » = le code devient la loi : la manière d'écrire le code va influencer la manière dont les règles seront appliquées, on a donc des présupposées.

→ smart contract : exécute les termes d'un contrat lorsque certaines conditions sont réunies. Tout est automatisé. C'est donc l'informaticien qui fait les règles et non pas des juristes, spécialistes du droit du contrat. S'il y a un bug, cela aura un impact sur le contrat.

Solution : le législateur utilise l'interdiction de la clause abusive, au cas par cas. Les concepteurs se sont aussi dit prêts à prévoir des règles spécifiques.

2. Delphine Harou

(secrétaire du groupe consultatif sur l'éthique lancé par le Contrôleur européen de la Protection des Données (CEPD) et membre de l'équipe de supervision et de mise en application du CEPD)

European Data Protection Supervisor = l'équivalent de la CNIL au niveau européen

Missions :

- Données de l'espace Schengen, Eurodac
- Faire remarquer les enjeux au Parlement et au Conseil
- Coopération avec les autorités/commissions nationales (comme la CNIL)
- Dimension éthique

Pourquoi s'intéresser à la dimension éthique du Big Data / algorithmes ?

- Nous sommes personnellement concernés car le carburant = nos données.
- Nos principes sont mis sous tension par le phénomène de Big Data / algorithmes.
- La révolution numérique a un impact collectif important.

→ Les enjeux sont mondiaux. Qu'est-ce qui caractérise la « révolution numérique » ?

- Masse énorme de données
- Capacité de stockage et de traitement des données
- Techniques d'analyse plus puissantes
- Objets connectés

Avec nos données, on laisse des traces partout.

Le Big data crée des tensions avec la protection des données personnelles :

Normalement, l'objectif devrait être la minimisation des données, mais aujourd'hui on semble loin de cela. En effet, la collecte des données doit se limiter à une certaine finalité. Or, aujourd'hui les données sont là en abondance, ce qui permet de faire des corrélations entre ces données.

Il y a également un problème concernant l'anonymat des données personnelles :

Il ne faut pas que les données permettent d'identifier la personne. Mais l'anonymisation est de plus en plus impossible à cause de l'évolution des techniques, et ce qui est anonyme aujourd'hui ne le sera peut être pas demain.

La dimension éthique concerne les conséquences de ces évolutions numériques sur l'individu et la société dans sa totalité. Exemples :

- La **donnée comme monnaie d'échange** contre un service (Facebook, assurances...)
- **Fin de la mutualisation des risques** : On identifie chaque personne, qui va payer en fonction du risque qu'elle pose. Si on connaît les données individuelles des personnes, on peut déterminer précisément le montant d'une assurance par exemple. Les personnes vulnérables risquent d'être encore plus vulnérables, et cela pose un problème de discrimination et de justice sociale.
- **Discrimination** : le data set qu'on injecte à l'algorithme contient déjà un biais (par exemple discrimination homme/femme), qui va se multiplier à la sortie. Le problème : comment contester un algorithme ? Comment savoir que les gens sont discriminés ? La discrimination = invisible.
- **Perte de l'autonomie des individus** : On accepte certaines choses sur Internet qu'on n'accepterait pas dans le monde réel. Par exemple, on n'accepterait pas d'être suivi dans la rue, d'être abordé par des gens qui veulent tout le temps nous influencer, nous dire pour qui voter.
 - Moins d'autonomie dans la prise de décision en général
 - Pourquoi l'accepte-t-on ? Il y a un décalage entre le coût et les bénéfices qui n'est pas visible : les bénéfices sont directs : on peut directement avoir accès à un service tandis que les coûts ne sont pas visibles, ils se montreront à long terme et pas de façon explicite.

La question : laisse-t-on les technologies nous guider ou peut-on agir avant ?

Règlement européen d'application des données, applicable en 2018 :

- prévoit le droit à l'oubli, le déferencement, la portabilité.
- Nouveau principe de responsabilité = accountability, avec l'objectif d'un cercle vertueux de responsabilisation.
- Fairness/loyauté dans la façon dont on traite les données
- Transparence
 - o envers les utilisateurs (quels objectifs, la pondération des critères...)
 - o envers les autorités de contrôle

Les différences culturelles concernant l'éthique et le numérique :

- USA : il n'y a pas de loi sur la protection des données → vide juridique. Utilisation d'autres droits, comme la protection du consommateur par exemple.

Il y a une sorte d'auto régulation : les associations professionnelles vont se mettre ensemble et chercher qu'est-ce que l'éthique par rapport aux utilisateurs → « algorithmic fairness ». Ils traduisent les valeurs en terme de principes que les programmeurs vont pouvoir utiliser.

- Union Européenne : avant tout basé sur le droit et la protection des données. Il faut donc rajouter la dimension éthique et mettre l'humain au centre de la réflexion.

Le droit de ce domaine est très technique. Il faut donc éclairer l'esprit de la loi avec ces valeurs, avec l'éthique.

Quelles alternatives seraient possibles ?

Payer d'une façon différente qu'avec les données, par exemple sur Facebook? Car si on continue de monnayer les données, ce sont les plus pauvres qui vont avoir besoin de vendre leurs données, et les données des plus pauvres ont moins de valeur car ils ont moins de pouvoir d'achat → poursuit les

discriminations. Il ne faut donc pas entrer dans la monétisation des données personnelles, plus que ce qui existe aujourd'hui.

Pourtant, il ne faut pas oublier les bonnes choses du Big Data et les opportunités qui existent.

Selon Madame Harou, on peut comparer ces questions aux enjeux de l'écologie il y a dix ans : aujourd'hui, ils sont entrés dans les mœurs, mais ça a commencé avec des petits actes au quotidien que l'on devait rappeler aux individus. Ainsi, au niveau individuel, on pourrait faire des petits actes qui ne sont pas révolutionnaires mais presque, comme ne pas utiliser Google ou ne pas être géolocalisé en permanence.

3. Ludovic Van Egroo

(enseignant, consultant Organisation Altedia Lee Hecht Harrison, spécialisé en Due Diligence et Open Source)

Exemple des algorithmes dans le domaine des ressources humaines.

Les algorithmes pourront être utilisés pour optimiser la gestion quotidienne de l'entreprise : anticiper, comprendre le fonctionnement humain de l'entreprise, par exemple pour anticiper les absences, en connaître les causes possibles ; mais aussi recruter le meilleur profil en anticipant l'avenir du secteur (cf. beaucoup d'entreprises n'avaient pas anticipé le besoin de compétences digitales). → Permet d'aider à la prise de décision des dirigeants, et de gagner en efficacité.

Cela passe par une préparation des données, une anonymisation des données pour ensuite en faire une analyse statistique et prédictive.

On a de plus en plus de données (Big Data) → les traiter devient humainement impossible, car complexe et chronophage. → Solution : intégrer algorithmes/intelligence artificielle et outil de traitement des données.

Nécessité de pluralité des sources :

- Collecte en interne (constat)
- Collecte en externe (open data du gouvernement et open source)

Une des idées : développer un outil pour consolider l'ensemble des données disponibles, paramétrer l'analyse des données et mettre en perspective avec la conjoncture actuelle, anticiper la prise de décision.

4. Eléments de débat

- « Privacy by design » : penser la protection des données dès la conception.
il faut aussi vérifier que tous les sous-traitants se conforment aux règles
- Le problème des textes juridiques : ils ne sont pas transversaux, il y en a beaucoup en parallèle ce qui rend le domaine encore moins transparent
- Le texte de l'Union européenne est bien, il y a déjà beaucoup d'entreprises, notamment américaines, qui anticipent son entrée en vigueur
- Différence : Google – pas payer au début, on « paie avec des données » / Apple – cher au début, on paie le produit, mais protection des données
- Les données d'une seule personne n'ont aucun intérêt, c'est la masse de données qui fait la différence
- Quels coûts ? Perte d'autonomie, transformation de la société/civilisation
- Blockchain : il faut faire attention aux discours marketing
- Protection des données peut devenir une arme entre concurrents

- Droit à l'oubli : privacy/ data protection by design est la seule possibilité → il faut penser différemment → Ce n'est pas parce que c'est trouvable sur Internet qu'on peut librement l'utiliser
- Exemple de la Chine : le gouvernement veut mettre en place un système de « social scoring » : récolter toutes les données de chacun et donner une note
- Notation déjà omniprésente : hôtels, restaurants, chauffeurs Uber
- Le système de collecte des données nous amène à mentir (sur notre âge, nom...), à éviter de laisser les bonnes traces → mais quel est l'impact sur notre société si manipuler / mentir devient indispensable?
- Conséquences pour la démocratie : élections, problèmes de faux robots qui manipulent
- Est-ce que des règles sont un frein pour le développement ? Plutôt non parce que les lois nécessitent la confrontation approfondie avec ce que l'on fait → il faut d'abord savoir ce qui se passe « à l'intérieur de la maison » avant d'agir à l'extérieur
- Question de la conservation des données : aujourd'hui selon la devise : « si jamais »/ « un jour, ça pourrait servir », donc quasiment tout est gardé. Or : la seule raison pour garder devrait être une finalité précise.