

COMPTE RENDU

Événement de lancement du cycle de débats publics sur les enjeux éthiques des algorithmes

Organisé à la CNIL, le 23 janvier 2017.

Allocution d'Isabelle FALQUE-PIERROTIN (Présidente de la CNIL)

La loi pour une République numérique a chargé la Cnil d'organiser la réflexion sur les enjeux de société et d'éthique liés au numérique. Cette mission sera honorée en constituant un débat public sur le sujet des algorithmes. Le choix de ceux-ci se justifie car ils sont au cœur de la vie numérique et concernent quotidiennement tous les secteurs de celle-ci : culture, finance etc... Porteurs de promesses, les algorithmes font cependant peur aux français. En effet, comment se construisent ces promesses ? Sont-elles maîtrisées, contrôlées ? C'est tout l'objectif de ce débat que de lever le voile sur ces questions et d'aborder celles-ci dans les secteurs de la justice, de l'économie, la santé etc...

L'idée du débat public a été retenue, car l'éthique du numérique nous concerne tous. Ce processus est collectif : lancé par la Cnil, il sera animé par tous les partenaires, publics ou privés qui expriment le souhait d'y participer. Certains participants nous ont déjà rejoint (dix-sept) appartenant à des champs variés. Leur objectif est d'organiser, à leur tour, des manifestations sur les enjeux éthiques soulevés par les algorithmes. La cohérence de l'ensemble de ces manifestations sera orchestrée par la Cnil et une restitution du panorama produit sera faite pour l'automne 2017 à l'issue de laquelle nous serons en mesure de cartographier les arguments et les enjeux éthiques des algorithmes en fonction de la variété des champs abordés, pour en tirer des recommandations de politique publique.

Présentation sondage IFOP par Jérôme FOUQUET

Restitution des résultats d'une enquête récente sur le sujet des algorithmes.

8 français sur 10 disent en avoir déjà entendu parler. Seul 1/3 savent précisément de quoi cela relève et la moitié en ont connaissance sans pouvoir cependant en dire plus.

Sur ce genre d'enquête se reflètent des clivages classiques, socio-éducatifs ou socio-culturels. Les catégories CSP+ ont une connaissance assez précise du sujet pour 48% d'entre-elles, contre 22% seulement pour les CSP moindres (ouvriers/employés).

La deuxième question, après avoir posé une définition précise des algorithmes, consistait à demander aux Français interrogés si, à leur sens, ces outils étaient présents, ou non, dans leur vie de tous les jours : 80% répondent oui, avec plus d'un tiers qui pensent qu'ils sont très présents ; 43% pensent qu'ils sont assez présents et 1/5 qu'ils ne sont pas présents dans leur environnement quotidien. L'écart de réponse entre les CSP supérieures et inférieures s'étant réduit après définition.

8 français sur 10 nous disent que ces outils/technologies sont présents dans nos vies quotidiennes, 2/3 d'entre eux pensent que nous ne sommes qu'au début de ce processus et qu'ils seront appelés à avoir une place encore plus grande.

S'agit-il d'un enjeu de société (au sens citoyen du terme) ? Oui pour 3/4 des personnes interrogées, avec une opinion molle de 54% des personnes pour qui il s'agit plutôt d'un enjeu de société.

Les questions suivantes sont binaires : est-ce que les algorithmes sont fiables ? 47% répondent plutôt fiables, 53% plutôt sources d'erreurs. C'est à relier au fait que le niveau de connaissance pour avoir une opinion précise de cette technologie est parcellaire, ainsi qu'au contexte, ambivalent dans l'opinion à l'encontre des technologies en général.

Est-ce que les algorithmes proposent plus de choix aux individus grâce à une meilleure connaissance de leurs comportements et de leurs pratiques ? 43% répondent favorablement, contre 57% qui estiment que les choix deviennent limités, imposés ou répétitifs.

Est-ce que les algorithmes représentent des opportunités pour chacun grâce à l'accès à des services plus personnalisés ? Seul 1/3 répondent favorablement, contre 2/3 qui pensent qu'ils représentent plutôt une menace en raison de l'accumulation des données personnelles sur les choix, les goûts et les comportements de chacun.

L'interprétation de ces résultats est liée au contexte national et culturel de chaque pays et sera différent en Allemagne par exemple.

La variable générationnelle est perceptible, plus le public interrogé est jeune, plus il est bienveillant ou positif vis-à-vis de cette technologie. Par ailleurs, les CSP supérieures se montrent moins inquiètes, moins négatives quant à l'utilisation des algorithmes

1 14h30 - Table-ronde : des algorithmes et des hommes

Présentation du programme de la première table ronde par Edouard GEFFRAY (EG).

1.1 Algorithmes, intelligence artificielle : quelles définitions ?

EG : présentation des invités, Paul DUAN (**PD**), Rand HINDI (**RH**), Roger-François GAUTHIER (**RFG**), Jean-Philippe DESBIOLLES (**JPD**) et Antoine GARAPON (**AG**).

Au vu du sondage présenté peu, de personnes savent précisément ce qu'est un algorithme. Quelle serait la définition technique d'un algorithme ?

PD : prendre une bonne décision nécessite d'avoir un maximum d'informations mais on génère aujourd'hui une quantité exponentielle de données que le cerveau humain n'est pas en capacité de traiter manuellement. On a donc inventé des formules mathématiques, des outils pour le faire à notre place : l'utilisation des algorithmes est donc d'abord une réponse à ce problème.

Il faut ensuite distinguer Intelligence artificielle (IA) forte et Intelligence artificielle faible : une IA faible, ce sont des algorithmes extrêmement spécialisés pour effectuer des tâches régulières, préprogrammées, pour résoudre des problèmes précis. Une IA forte serait une IA capable de résoudre des problèmes généraux.

RH : il existe un amalgame entre les algorithmes et l'IA. Certaines définitions données sont proches du *machine learning*.

Or un algorithme n'est d'abord rien d'autre qu'une procédure. C'est une liste de règles qui peuvent être apprises à partir de données ou écrites à la main par des humains et qui sont ensuite appliquées de manière automatisée. Le *machine learning* n'est qu'une petite partie de l'algorithmie, mais il en constitue une partie croissante aujourd'hui. Avec les algorithmes de *machine learning*, au contraire, la question des règles sous-jacentes n'est plus posée. L'un des avantages principaux c'est que cela permet d'aller très vite, sans avoir en plus la nécessité de comprendre le phénomène pour être capable de reproduire un comportement.

JPD : effectivement, on passe aujourd'hui d'un monde de programmation à un monde d'apprentissage. La programmation, l'exécution existe depuis près d'un siècle, c'est une réponse à un besoin exprimé, le programme exécute ce pour quoi il a été conçu. On se dirige désormais vers un monde d'apprentissage, on apprend à un système à raisonner, à penser, qu'est-ce qui est juste, qu'est-ce qui ne l'est pas etc...

On assiste également à un transfert d'expertise : les experts, les sachants, passent aujourd'hui du temps avec les machines pour leur transmettre leur savoir. La machine elle-même partagera ensuite avec un grand nombre ce qu'elle a accumulé en ayant été formée par un petit nombre.

Un système cognitif, tels que ceux avec lesquels nous travaillons, a aujourd'hui trois caractéristiques :

- il dialogue en langage naturel avec l'être humain (NLP : Natural Language Processing).
- La deuxième caractéristique c'est que ce sont des systèmes probabilistes et non pas déterministes.
- La troisième caractéristique c'est que ce sont des systèmes apprenants, avec une logique de *machine learning* et de *deep learning*, chaque fois que le système est utilisé, il apprend et s'améliore, statistiquement parlant, en fonction de l'utilisation qui en est faite.

RFG : à propos du sondage : si l'on peut comprendre que les jeunes générations soient moins inquiétées par l'essor des algorithmes, j'espère cependant qu'elles sont critiques vis-à-vis d'eux. La connaissance de ce qu'est un algorithme et de ses enjeux devrait être enseignée à l'école. L'objectif n'est pas d'inquiéter les gens, mais de leur donner des éléments d'analyse. La connaissance par les gens des enjeux des algorithmes est aujourd'hui fondamentale.

PD : il est également intéressant de mettre en avant le contexte dans lequel a été réalisé le sondage, on pourrait se demander si l'opinion que les gens ont sur les algorithmes aurait été différente avant ou après l'élection de Trump par exemple.

AG : à la fin de l'exposition Rembrandt, on peut voir un tableau peint par une IA à la manière de Rembrandt. Quelle différence entre le dessin de l'IA et le dessin de Rembrandt ? Cela pose la question de l'individuation de la forme, du dessin, de l'intention humaine. Dans un cas, quelqu'un nous raconte quelque chose de la vie, de lui-même, de nous et dans l'autre on a un vrai tableau, mais qu'est-ce que ce tableau ? C'est un objet qui existe, car fabriqué, mais qui n'existe pas en tant qu'œuvre d'art... C'est à méditer.

EG : Rand, vous avez dit qu'un algorithme, cela prend des décisions sans qu'il soit besoin de se poser la question des règles sous-jacentes. Comment fait-on alors pour exercer son esprit critique, si l'on n'a pas besoin de se poser la question des règles sous-jacentes? A quel moment réinterroge-t-on la logique de la machine?

RH : la question ne tourne pas autour d'une IA qui aurait une capacité complètement autonome dans un monde complexe. Ce n'est pas parce qu'elles sont capables de copier le style d'un artiste, qu'elles seront capables de capturer l'essence du monde et de l'exprimer sur un support, à travers un dessin.

Si l'on voit les IA et les algorithmes comme étant des outils qui peuvent apprendre à effectuer des tâches spécifiques, mieux que les humains et de manière plus fiable, statistiquement on peut leur faire confiance. Exemple avec les voitures autonomes. Par contre, les erreurs que font les algorithmes et les humains ne sont pas les mêmes: une IA est plus fiable et fait moins d'erreurs, statistiquement parlant. Les IA font moins d'erreur que les humains mais elles font des erreurs là où des humains n'en auraient pas fait. C'est ce qui est arrivé avec l'accident de la voiture autonome de Tesla, qui ne serait jamais arrivé avec un humain.

PD : Que veut dire « être bon » ou « être fiable » pour un algorithme ? Le résultat d'un algorithme est fonction de son contexte d'utilisation. Son emploi n'est pas obligatoirement neutre et objectif. Exemple avec l'algorithme de recommandation Youtube, finalement assez simple, dont la fonction est de mettre en avant la vidéo qui retiendra le plus longtemps le visiteur. On voit ici qu'on aurait pu privilégier une autre métrique et que les éventuels effets pervers et biais sont liés au choix de la métrique au moment de la conception. Il est important de veiller à ce que l'emploi des algorithmes n'ait pas d'effets pervers.

Exemple également dans le domaine de la récidive aux Etats-Unis, avec des produits comme COMPAS ou le modèle LSER, qui en analysant certains facteurs permettent de prédire si un condamné va récidiver. Au-delà du glissement dans l'utilisation de l'algorithme, initialement prédictif mais qui a fini par être employé par les juges pour prendre des décisions, on a découvert qu'alors qu'il n'y avait pas de variable race dans l'analyse des données, les résultats étaient néanmoins plus à charge contre les gens de couleur.

Il est donc important de comprendre l'algorithme et de mettre en valeur son *explicabilité*. On doit pouvoir comprendre son raisonnement.

JPD : il existe trois notions très différentes les unes des autres qu'il faut distinguer :

- les systèmes d'IA cognitive (qui fonctionnent avec des informations non structurées),
- les systèmes prédictifs, qui obéissent à une notion d'analytique et de données structurées, avec des *patterns*, de suites arithmétiques etc...
- et les systèmes prescriptifs, en vue d'une recommandation à quelqu'un et d'une prise de décision, d'une action.

Chacune de ces trois notions répond à des logiques très différentes, des outils différents, des compétences différentes et ne sont pas obligatoirement liées les unes aux autres.

Sur la prise de décision, les systèmes cognitifs, systèmes que les anglo-saxons appellent les *evidence based systems*, ne se substituent pas aux docteurs et au corps médical, l'objectif est d'assister et d'augmenter. (on préfère ainsi parler d'*augmented intelligence* plutôt que d'*artificial intelligence*). Ces systèmes exposent les faits qui les ont amenés à la recommandation, c'est inclus dans leur design même. Cela augmente en fait la traçabilité de la prise de décision ! Avec ces systèmes, on est aujourd'hui en mesure de dire *pourquoi* il vaut mieux faire A que B.

AG : pour en revenir à l'*evidence base sentencing* en matière de justice, la difficulté n'est pas que les prédictions soient fausses, c'est qu'elles aillent à l'encontre de principes constitutionnels. Pour un délit identique, un blanc aura trois mois de prison avec sursis, tandis qu'un noir cinq ans d'enfermement. La fiabilité des algorithmes est à débattre.

RH : que cherche-t-on à optimiser avec ces algorithmes? Ce ne sera pas la même chose pour Youtube et pour la justice ou la médecine. La métrique essentielle ne sera pas la même. Pour Youtube, la métrique, le critère utilisé, sera celle qui vise à ce que l'internaute passe le plus de temps sur la plateforme. Dans le domaine médical, par exemple, on espère que la métrique choisie n'est pas celle qui vise prioritairement à optimiser l'occupation de l'hôpital. Tout cela est donc moins une question d'algorithmie que de design, de priorité pour un business.

PD : Toute métrique est toujours imparfaite. Il est crucial de bien définir les paramètres choisis. Cela pose la question du cadre, qui même s'il est clair au départ, n'est jamais définitif et peut éventuellement intégrer des biais qui ne sont découverts que plus tard.

RH : on peut biaiser l'apprentissage d'un algorithme en sélectionnant mal les données qui lui seront apportées. Il faut donc faire attention à la fois aux données qu'on entre dans un système et aux métriques qu'on choisit. On peut imaginer que dans des domaines sensibles comme la justice ou la santé, des commissions pourraient auditer à la fois les données entrantes mais aussi les mesures de performance pour éviter des biais ou une malveillance.

JPD : dans les systèmes cognitifs, il faut définir :

- le *knowledge*, le corpus, les données qu'on entre dans le système
- le système d'apprentissage et la méthode de raisonnement, le système de valeur (bien/mal, juste/injuste) qu'il intègre

AG : il ne faut pas se laisser berner par l'illusion de la performance potentielle, par les usages. Le problème, c'est que l'emploi des algorithmes développe une forme de délégation implicite à une nouvelle caste de scribes, seule à même de comprendre et de maîtriser les algorithmes, alors qu'il a fallu des siècles pour que le savoir soit justement distribué et réparti. Cela constitue une forme de retour en arrière.

JPD : il est intéressant que la connaissance développée par un médecin après 25 ans de travail sur une pathologie très spécifique soit transmise au plus grand nombre. C'est un processus possible grâce aux algorithmes.

RH : ce n'est plus tant l'algorithme qui compte, il finira par être accessible ou révélé à tous, ce qui compte c'est la donnée utilisée pour faire apprendre les algorithmes autant que l'accès au marché des usages qui vont en être faits. Et cette donnée n'est pas obligatoirement entre les mains de ceux qui développent des algorithmes. Ainsi, contrairement aux scribes qui étaient les seuls à posséder la technique de l'écriture, il faut également le pendant inconditionnel à cette technique que sont les données pour les algorithmes.

EG : différents éléments émergent: les trois types de rapport au monde et à la connaissance (cognitif, prédictif et prescriptif) ; mais également la question de savoir comment incorporer l'esprit critique, un contrôle de leurs règles de fonctionnement afin d'éviter le piège de l'algorithme biaisé, l'endoctrinement de l'algorithme ; et enfin dans le système de chaîne des valeurs, la différence entre la donnée, l'algorithme et l'accès au marché.

Question de Laurence El Khouri (CNRS) : Au vu de la grande masse de données que seuls les algorithmes peuvent traiter, comment vérifie-t-on les calculs? Ou que les données utilisées sont les bonnes, si par définition l'homme ne peut combiner une telle somme de données? Peut-on se fier à la machine?

JPD : les systèmes cognitifs qui travaillent sur des données non structurées s'appuient sur une compréhension linguistique. L'objectif c'est de travailler sur les données non structurées qui représentent 80% des données disponibles autour de nous, comme des inscriptions tirées d'un carnet de notes par exemple, le système comprend le sens, l'intention de ce qui est écrit.

Le prédictif quant à lui, répond à des données structurées, c'est une suite mathématique, via des *patterns*, on analyse le passé pour prévoir l'avenir. Enfin, le prescriptif, qui donne une recommandation sur ce qui vient d'être dit.

Dans le cadre médical, chez nous, le médecin peut ou pas valider la recommandation, le système lui ayant produit les évidences de son raisonnement. On sait pourquoi le système fait telle ou telle recommandation et les faits sont consultables. Ce type de système vous apporte en fait l'auditabilité qui sous-tend la recommandation.

Question : certes, s'il s'agit de text mining. Mais s'il s'agit de mathématiques, comment vérifie-t-on ?

RH : comment fait-on pour détecter les biais dans les grands volumes de données? En programmation, on sait depuis longtemps faire des tests pour vérifier les programmes. Le même principe pourrait également être appliqué au *machine learning* pour tester les modèles de comportements qui ont été appris, avec des données vraies ou fausses qui permettent de tester les résultats de l'apprentissage. La mise en place de tests de *machine learning* va se mettre en place progressivement dans un avenir proche.

Par ailleurs, la plupart des erreurs surviennent car la qualité de la donnée entrée était mauvaise. Aujourd'hui la plupart des algorithmes apprennent à partir de données de « white american guys ». Les algorithmes sont donc biaisés vers ces catégories de données. Les données entrées reflètent un état d'esprit de société dont les algorithmes ne sont que le miroir. On travaille par ailleurs à éviter ces biais pour augmenter la fiabilité des systèmes. Il faut mettre en place des méthodologies pour remédier à ces biais inscrits dans les bases de données mêmes. Mais personne n'a en fait encore de réponse.

Question : Si ce qui a de la valeur c'est la donnée, comment exploite-t-on cette richesse ? A qui appartient-elle ? Peut-on imaginer qu'un individu soit rétribué pécuniairement sur les données qu'il génère ?

Question de Claude Kirchner (Inria) : Attention à ne pas biaiser le débat en le centrant sur le *machine learning*. La question ne tourne pas exclusivement autour du *machine learning*, les questions d'éthiques et de compréhension intéressent aussi des algorithmes plus simples qui marquent déjà notre quotidien : ceux utilisés pour faire marcher une calculatrice (pour calculer 2+3), ou encore Admission Post Bac (pour assigner une université aux étudiants qui le demandent), par exemple. Or, ces algorithmes-là sont déterministes, on peut démontrer leur correction, le fait qu'ils suivent l'intention de leur concepteur. La réflexion éthique sur les algorithmes doit nécessairement prendre en compte ces algorithmes-là. Enfin, seulement, on peut étendre le débat aux algorithmes capables de s'automodifier, avec le *machine learning* (algorithmes qui peuvent être utilisés dans ce qu'on appelle l'IA). Mais il ne faut pas se focaliser sur les aspects « *machine learning* » car on risque d'oublier toute une classe d'algorithmes, plus simples, mais d'ores et déjà très présents. Ensuite on peut aller vers les algorithmes de *machine learning*, plus difficiles à comprendre.

RH : la question de la propriété de la donnée que l'on fournit à l'IA est une question indépendante de l'IA. On peut imaginer qu'elles n'appartiennent pas exclusivement à un individu et qu'il y ait même des licences d'exploitations entre entités détentrices ou génératrices de données. Tout porte à croire qu'il serait positif que les données appartiennent aux individus, ne serait-ce que pour éviter un monopole des GAFA et favoriser la concurrence.

EG : on note dans les débats que l'algorithme apparaît comme secondaire par rapport à la donnée. Ainsi, lorsqu'on parle d'éthique des algorithmes, peut-on parler des conditions éthiques pour accéder à des gisements de données ?

RH : oui, comme avec des matières premières dans le monde réel. La donnée est à l'Intelligence Artificielle ce que l'uranium est au nucléaire. Avec la même problématique des usages qui en sont faits (civil ou militaire).

1.2 Quelle place de l'homme dans les systèmes algorithmiques ?

EG : nous avons évoqué la place de l'homme dans le monde des algorithmes, notamment pour ce qui concerne les contrôles. Mais l'algorithme est-il un substitut, un prolongement, une aide à la décision humaine, un gain de productivité ?

AG : il est important que l'homme garde sa place de décideur. Il y aurait un grand risque de performativité, qui troublerait la décision, dans le monde juridique. Il y a un rêve d'automaticité avec l'IA, qui est impossible dans certains secteurs. Dans le fonctionnement de la justice, l'humain est fondamental avec ce qu'il implique d'interprétation des règles et de débats. Il faut toujours revenir à l'humain.

Nous serons bientôt confrontés aux *smart contracts*, qui n'auront plus d'expression langagière, de langage ordinaire mais qui n'auront qu'une expression algorithmique, à l'exécution automatisée. La *blockchain* par exemple apporte une fiabilité inégalée par les institutions humaines. Mais tout l'aspect interprétatif, des textes, des faits, risque de disparaître. Alors qu'il fait partie intégrante de l'univers juridique. Cela pose des questions. Qui prend la décision et quelle est la place de l'homme dans la justice prédictive ?

EG : dans les domaines régaliens et dans celui de l'Education Nationale en particulier, le pouvoir se définit en principe par la capacité de décision. Aujourd'hui, moins on apprécie le pouvoir discrétionnaire et plus on se fie aux algorithmes. Comment justifie-t-on cette évolution et quelle place reste-t-il à l'Homme ?

RFG : l'humain est aussi bien en amont qu'en aval de la décision. Les algorithmes très connus, type Admission Post Bac ou bien AFFELNET ont pour objectif la répartition d'un grand nombre d'élèves et d'étudiants. Ils ont fait quelque chose de remarquable : ils ont mis fin à un fonctionnement mafieux. Auparavant, ces décisions de répartition se prenaient dans le secret des bureaux des proviseurs et des inspecteurs d'académie avec des piles de recommandations.

Des travaux de sociologues, qui ont observé le fonctionnement réel d'ABP (Agnès VAN ZANTEN par exemple), marquent bien l'importance des données. Ces systèmes font appel à des vœux, hors tout le monde souhaite le meilleur. Mais il existe une inégalité extrême dans la population des lycéens, entre ceux qui sont capables de formuler des vœux, car ils sont capables d'élaborer des stratégies que d'autres ne sont pas en mesure d'élaborer. L'expression des vœux n'est pas neutre en soi.

Par ailleurs, l'enseignement privé, quelques grands lycées n'entrent pas dans ce système automatique. C'est précisément dans l'espace qui échappe au fonctionnement automatisé par l'algorithme qu'une nouvelle forme de mafia va s'engouffrer : le principe de création de ce système automatisé à vocation objective est louable et vertueux, mais comme il ne s'applique pas à l'ensemble, on maintient un déséquilibre et une situation liée à l'arbitraire.

Un autre pendant négatif à ce système, c'est qu'il devient plus important que l'enseignement même. On le constate dans les derniers mois dans les classes de terminale par exemple. Cette hiérarchisation des critères de sélection, des classements des lycées ne fait que renforcer l'idée que ce n'est pas l'enseignement qui compte, mais le lycée, collège ou université où l'on va être. Ce biais a été découvert plus tard, qui n'était pas l'intention de départ, l'objectif premier était de simplifier une tâche administrative.

Par ailleurs, un algorithme, dans l'Education Nationale, ce n'est pas nouveau : un conseil de classe qui décide quelque chose pour un élève à partir du calcul de la moyenne générale, que fait-il ? Le travail d'un algorithme. Donc je ne comprends pas l'utilité de remplacer un algorithme par un autre, si c'est pour en plus voir apparaître d'autres effets secondaires indésirables. Cette sacralisation de la note, dont on connaît les défauts (subjectivité, relativité etc...), comme donnée à entrer dans un algorithme est biaisée en soi, la computation de l'algorithme ne peut en ressortir que biaisée également.

PD : la question posée est aussi celle de savoir dans quelle mesure nous sommes prêt à troquer un biais humain contre un biais systématique. Les juges seraient par exemple plus sévères avant d'avoir mangé qu'après : avec les algorithmes, on cherche à retirer une partie de ce type de biais humain, mais la solution n'est pas absolue et de nouveaux biais apparaissent. Il est essentiel de prendre en compte l'utilisation et le contexte dans lequel va être utilisé l'algorithme, en incluant ces questions dans son design même. Ce n'est pas une science, mais un art.

L'humain a une place à tous les niveaux du processus algorithmique :

- Dans la génération des données (par exemple, certains prisonniers changent de comportement une fois incarcérés : insérer cette catégorie de données dans le système ou ne pas le faire a une influence sur la recommandation obtenue *in fine*),
- dans la conception de l'algorithme,
- dans la prise de décision (entre aide à la décision, il n'y a qu'un pas dans les faits: les juges submergés par une grande quantité de travail ne prennent plus cette recommandation comme une donnée parmi toutes les autres, mais comme la donnée décisionnelle Dans la conduite automatisée par exemple, on pourrait définir différentes configurations allant d'une automatisation complète du pilotage à une automatisation partielle qui nécessite une prise de décision du conducteur dans certaines situations)
- Dans l'audit des algorithmes.

Pour chaque sujet de société impliqué par les algorithmes, il faudrait réfléchir à des systèmes d'audit et de tests permettant de vérifier les systèmes algorithmiques.

JPD : la place de l'humain repose également sur le fait de *superviser* les systèmes d'apprentissage pour que leur développement ne soit pas indépendant ni autonome, afin d'en limiter les biais potentiels et les manipulations. J'insiste sur l'importance du contrôle de l'apprentissage, pour qu'un algorithme ne finisse pas raciste, misogyne comme certains exemples l'ont montré. Chez IBM, nous travaillons avec des systèmes supervisés : on observe comment le système se comporte, on décide avec le client si on réjette l'apprentissage dans le système. Superviser, c'est accompagner le système pour l'améliorer, le contrôler, éviter qu'il ne se comporte mal.

Il n'y a pas plus humain que l'IA : le système est fait par des hommes, supervisé par des hommes, pour accompagner des hommes.

Question d'Alexei GRINBAUM (CERNA) : Je souligne le fait que l'homme va changer, évoluer au contact de l'algorithme, un médecin va avoir tendance à faire confiance à l'algorithme, peut-être allant jusqu'à se déresponsabiliser. Quelles sont les évolutions du cadre juridique, du côté de l'homme ?

Question de Laurence DEVILLERS (CERNA) : L'IA n'est pas très humaine, c'est un algorithme, c'est mis en place par des humains pour des humains, mais ce n'est pas du tout humain. Et par ailleurs, dans le *deep learning*, n'y a-t-il pas des parties non supervisées ? Effectivement, il y en a ! La présentation de l'apprentissage ne peut pas être aussi simpliste, il y a des étapes supervisées, d'autres non, différentes méthodes de travail existent, les propos se doivent d'être précis pour éviter les confusions.

Question de Max DAUCHET (CERNA) : Il n'est pas évident de faire en sorte que les gens se posent les bonnes questions et soient critiques au sujet des algorithmes.

Le monde de la recherche publique est très soucieux des questions de la place de l'homme et du devenir social des objets que l'on construit grâce à la technologie. Il faut mettre les professeurs et élèves, au centre de ces questions d'éthique pour développer les opinions de demain. Le monde de la recherche est sensible au fait que toutes ces questions soient également liées à la pluralité.

EG : effectivement, la question de la pluralité est lancée au travers des partenaires de ce débat public qui organiseront des événements secteur par secteur, avec des objectifs et des niveaux de discussions différents pour chacun d'entre eux, que l'on assemblera à la fin pour obtenir une vision à 360 degrés.

Question d'Isabelle Falque-Pierrotin : on voit bien à travers tout ce qui a été discuté que l'une des questions à approfondir est la notion d'éthique de procédure (les algorithmes répondent-ils à nos valeurs dans leur fonctionnement ? Quels sont les biais éventuels ? etc...). Mais qu'en est-il de l'éthique de fond ? Existe-t-il des tabous (de fond) qui devraient être fixés dans le fonctionnement des algorithmes ?

AG : pour répondre à Alexei, la justice prédictive s'applique rarement, mais elle modifie considérablement l'attitude des acteurs, les poussant à négocier et à transiger. Dans le monde des avocats, certains clients ne souhaitent pas une justice bonne et rapide, mais au contraire, visent une procédure stratégique pour gagner du temps etc... En médecine, on a le même problème avec les patients qui ne veulent pas guérir. Concernant l'éthique de fond, il est en train de se constituer un comité d'éthique de la justice prédictive, réunissant des personnes qui font de la justice prédictive pour discuter des problèmes éthique qu'ils rencontrent. Cette justice prédictive constitue une radiographie de l'activité des juges, les mettant parfois face à leurs contradictions ou à leur production, dont eux-mêmes n'avaient pas conscience. L'analyse des statistiques de décisions d'un juge avec 30 ans de carrière peuvent mettre en avant une constante qui permettrait de prédire les probabilités de décision, créant une sorte de partialité avant même que la procédure soit engagée. Ce qui est nouveau avec les algorithmes de justice prédictive, c'est que le juge voit sa propre production ! Et cela peut poser des problèmes pour la crédibilité des institutions publiques.

PD : Il faudrait voir secteur par secteur quelle pourrait être la charte éthique à adopter. Parler de tabous, c'est se poser la question des *valeurs* que l'on veut promouvoir à travers les algorithmes aussi bien dans leur conception, dans leur contrôle que dans leur utilisation. Exemple : pour l'algorithme de Youtube, l'objectif pourrait ne pas être économique, mais celui de promouvoir la véracité de l'information. Quand on utilise des algorithmes dans le secteur de l'emploi, on encode en fait une certaine vision de ce qu'est le marché du travail, de ce qu'est la valeur-travail. Tout cela pose la question de qui décide de ces valeurs à inculquer et donc du rôle des institutions publiques.

RH : c'est effectivement une question politique, ce n'est pas parce qu'une partie des emplois peuvent être automatisés par des IA, que cela veut dire qu'ils seront automatisés ou que l'opinion publique acceptera qu'ils soient automatisés.

L'Homme ne s'oppose pas nécessairement à la machine lorsqu'on parle d'éthique. C'est le cas dans certaines situations, où la place de l'homme est indispensable, mais l'association de l'homme et de la machine est complémentaire dans d'autres. Leur relation n'est pas exclusive.

Une question centrale est celle de savoir quels sont les métiers qu'on se refuse à automatiser pour des raisons éthiques.

JPD : les éléments sont réunis pour déclencher une quatrième révolution industrielle qui devrait transformer les modèles structurels de communication, de travail etc... L'IA est un de ces éléments. Il est important de définir des chartes de valeurs qui seront applicables au monde de demain et avec lesquelles il est nécessaire de travailler maintenant : la transparence, l'augmentation et non pas le remplacement de l'homme. Effectivement cette question des valeurs est essentielle. Chez IBM, nous venons ainsi d'annoncer une charte des valeurs que nous allons appliquer à notre activité. Nous préférons parler de valeurs, plus que de tabous.

RFG : la signification éthique et politique des algorithmes est fondamentale. Dans l'éducation, avec les *learning analytics*, on peut faire de la prédiction qui pourrait déboucher sur un enfermement des élèves. Or, en France, il est probable que ce genre de problème suscite trop peu d'attention. Il faut pourtant faire en sorte que l'élève échappe au déterminisme et pour cela la question des valeurs inscrites dans les systèmes algorithmiques est fondamentale.

EG : quelle autonomie (au sens étymologique) dans un monde algorithmé ? Comment se fixer sa propre norme, être capable de l'exercer et de la conserver ? Un début de réponse se trouve dans la loyauté des plateformes et dans la transparence des algorithmes : c'est l'objet de notre deuxième table ronde.

2 16h15 -Table-ronde : loyauté, transparence et pluralité des algorithmes

Présentation du programme de la seconde table ronde par Joëlle FARCHY (JF)

JF : l'objectif de la seconde table ronde est d'ouvrir le débat en tentant de définir « l'algorithme optimal », non pas sur le plan technique, mais sur le plan sociétal, à travers les questions de loyauté et de transparence d'une part, et de pluralité, ainsi que son pendant inverse, l'enfermement, que pourraient créer les algorithmes d'autre part.

2.1 Loyauté, quelles sont les règles du jeu ? Existe-t-il un idéal de transparence ?

Avec une proposition de définition caricaturale, la loyauté serait que l'algorithme « fait ce qu'il dit » et la transparence serait « qu'il dit ce qu'il fait ».

La notion de transparence est complexe, elle regrouperait la capacité à voir un algorithme, comprendre comment il fonctionne, en plus d'être en mesure de le contrôler. Par ailleurs, cette même transparence est-elle un idéal à atteindre ? Peut-on éviter les conséquences négatives qui en découleraient : rupture du secret commercial des entreprises, difficulté d'interprétation pour l'internaute submergé par un flot d'informations supplémentaires... Pire encore, cette transparence, poussée à son absolu, ne serait-elle pas totalitaire, remettant en cause les règles de la démocratie ?

Dominique CARDON (DC) : la question de la transparence est effectivement complexe en soi. Avant cela, la question est de savoir : que doit-on rendre transparent dans l'algorithme ? Est-ce la technique statistique employée ? Faut-il rendre le code visible ? Même si c'est utile, il y a des raisons pour qu'il ne soit pas obligatoirement dévoilé. Exemple avec le marché du « search engine optimization » où des acteurs cherchent à tricher avec l'algorithme : cela permet de comprendre l'une des raisons pour lesquelles Google ne rend pas son code public.

Rendre transparent un calculateur, cela doit avant tout être un travail pédagogique, pour essayer de faire comprendre ce qu'il fait. Ce qui est essentiel, ce n'est pas que le code soit transparent, c'est que l'on comprenne ce qui rentre et ce qui sort de l'algorithme ainsi que son objectif. C'est cela qui doit être transparent.

Rendre le code transparent, ne veut pas dire qu'on va le comprendre, même par ceux qui le programment et notamment concernant les algorithmes qui sont en technique d'apprentissage massif.

De plus en plus, aujourd'hui, les règles de l'algorithme sont façonnées par les données que l'on fait entrer dans le système (on est de moins en moins dans un système organisé d'instructions) : on couple le système de règles programmées au départ avec les données qui sont analysées par l'apprentissage. Nous donnons des données pour les algorithmes apprennent.

On peut dès lors se poser la question de savoir si l'algorithme a bien appris ? A quelle école a-t-il été ? La réponse à cette question dépend de la qualité des données qu'on a demandé à l'algorithme d'apprendre.

Ces techniques d'apprentissage sont très opaques, autant que les données qui sont analysées, en plus d'être rares, à tel point que Facebook achète à des brokers publicitaires des données de segments commerciaux, sur nos consommations, pour les intégrer à leur calcul.

En conclusion, deux choses sont essentielles si l'on parle de transparence : savoir ce qui rentre dans la machine (quelles données ?) et quel est l'objectif d'apprentissage inculqué. Il faudrait rendre publics ces deux éléments.

JF : Antoinette ROUVROY (AR), comment s'articule la notion que vous avez développée de *gouvernementalité algorithmique* avec les objectifs de transparence et de loyauté ?

AR : le terme, un peu barbare, de *gouvernementalité algorithmique* a été choisi faute de mieux pour attester d'un glissement dans les manières de produire de la norme, à la faveur de la disponibilité massive de quantités de données numériques, qui induisent de nouveaux usages de la statistique.

Nous ne sommes plus dans une *gouvernementalité statistique*, qui présupposait des catégories, catégories préparées en fonction de conventions et qui préexistaient au traitement de données. Aujourd'hui, dans les

nouvelles pratiques impliquées dans les techniques d'analyse et d'exploitation des Big data, les catégories ne préexistent plus au recueil des données, ne gouvernant, ni ne limitant plus celui-ci. Au contraire, les catégories émergent après-coup du traitement de ces données, de manière spontanée, par la grâce d'algorithmes qui eux-mêmes apprennent du traitement des données et des faits, parfois sauvages, qu'ils rencontrent

On a l'impression d'une gouvernementalité immanente au réel numérisé lui-même, et qui fait que toute cette *gouvernementalité algorithmique* - en anglais, *data driven governance*, gouvernance par les données - apparait comme une forme révolutionnaire de gouvernement, dans la mesure où elle permet de s'émanciper du joug de la représentation, des institutions, de la norme déterminée par le Parlement, de la norme juridique etc... En vue d'une normativité beaucoup plus souple, sous des airs inoffensifs, avec l'apparence que le réel se gouverne lui-même, sans vision idéologique a priori.

Par exemple, avec les Big data et l'utilisation des algorithmes dans les campagnes électorales, on a l'impression que certains candidats pourraient vouloir détecter un programme électoral à même les données, ayant pour conséquence une émancipation illusoire pour l'électeur, celle d'être un élément du programme, chacun d'entre nous, de manière atomisée.

C'est une nouvelle manière de voir le monde et de gouverner, en structurant par avance les champs de pensées possibles des individus et donc de déconstruire les sujets, avec une hyper individualisation, une hyper segmentation autorisée par cette *gouvernementalité algorithmique*.

Par rapport à la loyauté, est-ce que les algorithmes font ce qu'ils disent ? C'est une partie de la question de la loyauté. L'autre partie concernerait l'évolution ou la disparition du pacte social, qui serait à établir aujourd'hui, entre les individus, les entreprises et l'Etat... voire même, peut-être entre les intelligences artificielles, à partir du moment où elles s'individualisent au fur et à mesure qu'elles apprennent.

De nouveaux acteurs, les GAFAs entre autres, ont acquis un pouvoir de privatisation de l'espace public tout à fait inédit. Cela concerne l'espace public actuel, mais aussi l'espace public potentiel : en effet, sachant que l'analyse de quantités massives de données permet d'ouvrir et de visualiser ce qui pourrait advenir, ce que les gens pourraient vouloir à l'avenir, toutes les opportunités et tous les risques futurs, c'est tout cet espace de potentialités qui se retrouve aujourd'hui accaparé par des puissances privées. Il y a une politique des espaces-temps potentiels à mettre en œuvre cet égard.

Concernant la transparence cette fois, je suis d'accord avec Dominique CARDON lorsqu'il dit que la seule transparence des logiques de traitement est insuffisante, car complexe pour la plupart d'entre nous, qui plus est difficile à transmettre : paradoxalement si l'on fait appel à des algorithmes, c'est parce que l'humain n'est pas en capacité de compiler une telle somme de données. Ces logiques ne sont donc pas transmissibles à un public non-averti, ni même transmissibles sous une forme linéaire ou linguistique.

De plus, la transparence des algorithmes aujourd'hui ne signifie rien sans la transparence des données. Or, celles-ci sont elles-mêmes rendues amnésiques de leurs conditions de production grâce aux exigences d'anonymisation posées par le régime de protection des données personnelles.

Il faut tendre autant que possible vers la transparence, mais celle-ci ne se suffit pas à elle-même, il ne faut pas s'en contenter. La transparence est insuffisante, surtout si elle dispense de la responsabilité. L'objectivité prêtée à la machine n'est pas encore l'équivalent de la justice. Que les algorithmes puissent être audités, dans leur design même, serait positif, mais là encore, insuffisant en soi.

JF à Bruno PATINO (BP) : dans le secteur des médias et de la culture, que vous évoque la notion de transparence et de loyauté ?

BP : Pendant longtemps je pensais que la transparence c'était d'aller regarder sous le capot et cela reste vrai dans le cadre d'une investigation journalistique. La réalité est plus complexe, le *machine learning* faisant que l'on a à faire à des animaux en mutation.

Ramenée à la culture et à l'information, la transparence des algorithmes, c'est d'avoir un contrat ou un pacte sur les effets qu'ils ont, c'est-à-dire, ce qu'ils se proposent de faire.

Dans la culture, tout est question de contextualisation, et c'est ce que font les algorithmes : ils contextualisent. Le contexte lui-même détermine la réception de ce que l'on va voir, lire ou écouter. Toute œuvre culturelle, ne peut être comprise et perçue que dans le contexte de sa réception, ce à quoi les algorithmes participent.

La transparence se joue donc aujourd'hui sur un contrat, une description préalable, de l'effet de contextualisation de l'algorithme : cet algorithme va contextualiser les choses de cette façon-là.

Exemple avec la télévision, qui jusqu'alors était contextualisée par des programmeurs, des humains qui se fondaient sur la rareté du temps et d'une grille : il y a 24 heures, 7 jours par semaine, on contextualise l'offre télévisuelle en fonction de cette rareté temporelle. Aujourd'hui, les plateformes comme Netflix, Amazon, ou autres, gèrent la problématique du choix, en combinant inflation et non-exhaustivité de l'offre culturelle par des algorithmes de recommandation. Ces algorithmes compensent la non-exhaustivité de ce qui est proposé par une contextualisation particulière. C'est dans ce cadre qu'il est absolument nécessaire de poser un principe de

transparence basique consistant à montrer « Voilà quels sont les grands principes de contextualisation de notre offre pour vous ».

En matière d'information cette fois, cette contextualisation pose problème puisque tout le monde ne reçoit pas forcément la même information, et même si c'était le cas, il la reçoit dans un contexte différent, ne la rendant donc pas identique pour tous. Cela pose la question de l'unicité de l'espace public qui se trouve dès lors fragmenté sous l'effet de la contextualisation et de l'individualisation. Or, Victor Hugo pensait que l'avènement d'une presse offrant à tous la même information était la condition de l'avènement de la République. Comment donc recréer un espace public derrière cette contextualisation et cette individualisation ?

Présentation de Transalgo par Nozha BOUJEMAA

Transalgo est né d'une mission du Conseil Général de l'Economie diligenté par Axelle LEMAIRE afin d'établir un rapport sur les modalités de régulation des algorithmes et de traitement de contenu. L'une des recommandations issue de ce rapport, rendu public le 15 décembre 2016, était la mise en place d'une plateforme scientifique collaborative destinée à favoriser le développement d'outils, logiciels et méthodes d'algorithmes et de promouvoir leur utilisation d'autre part.

Le rôle d'opérateur de cette plateforme, nommée Transalgo, a été confié à l'Inria avec le soutien du CNNum et l'IMT. Transalgo a pour objectif la transparence et la responsabilité des systèmes algorithmiques (ensemble algorithme et données) ainsi que des acteurs de l'économie numérique, avec un renforcement de la confiance envers le numérique et de développement du marché et des services numériques.

Certaines plateformes sur le marché jouent un rôle de prescripteurs en orientant une grande part du trafic, à travers les mécanismes de tri et les moteurs de recherche. Qu'en est-il de leur neutralité ? Qui peut la vérifier ? Cela soulève de nombreuses questions relatives à la neutralité, la loyauté, l'équité, la non-discrimination, la concurrence déloyale... Il est nécessaire de travailler collectivement sur les notions de transparence, responsabilité et explicabilité des algorithmes.

Exemple avec la recommandation de produit : est-ce que le produit suggéré est le plus approprié au besoin du consommateur ? Ou est-il fonction du stock du vendeur ?

Les moteurs de recherche peuvent ordonner leur classement sur des bases non-objectives, mais d'incitation commerciales. On sait que la question se pose chez certains industriels et que cela peut même constituer un business model innovant, l'essentiel c'est que le consommateur soit informé et qu'il ait le choix.

Les travaux de transparence représentatifs qu'étaient l'expérience Mobilities avec la Cnil et l'Inria, sur l'opacité de l'utilisation faite des données à caractère personnel des utilisateurs et sur leur consentement, soulevait déjà la question des informations entrantes et sortantes. Avec des logiciels capables de mesurer la quantité d'informations entrantes, sortantes et les appels système.

Il faut développer des outils algorithmiques capables d'effectuer ce genre de mesures, car ils ont prouvé qu'ils permettent de remonter jusqu'au responsable de traitement pour le mettre face à ses contradictions et à ses choix de traitement de données personnelles.

Par exemple, l'application Lampe Torche pour être installée nécessite d'avoir accès à nos contacts... Il faut se demander pourquoi ?

L'objectif est de répondre à toutes ces questions, en mettant en place des outils eux aussi algorithmiques, afin de mesurer la transparence des données et des algorithmes.

Au-delà de ça, à qui s'applique la loyauté ? Au consommateur, ou au fabricant du logiciel ? Cela dépend du contexte d'usage.

La transparence consiste à étudier les comportements d'un algorithme et à en vérifier la conformité à la loi et aux règles éthiques. Nous n'avons pas besoin de connaître le code source.

J'insiste enfin sur le besoin de développer des algorithmes responsables par construction, il faut dès le départ prévoir l'auditabilité, la transparence etc... Qui facilite l'explication et la traçabilité de leur raisonnement.

Intervention de Jean-Luc Molins, Secrétaire National UGICT CGT (CGT Cadres): Une contribution au débat, par rapport à la relation au travail, à l'éthique et à l'utilisation des données des salariés. celles-ci devraient selon moi être analysées sur trois niveaux: la collecte des données, que peut-on en faire et enfin où sont-elles stockées ?

Cette question de l'éthique au travail est essentielle à mes yeux, selon un barème récent avec l'institut ViaVoice, 57% des cadres rencontreraient un problème d'éthique dans leur activité professionnelle. Si on leur pose la

question de savoir s'il souhaiteraient disposer d'un droit d'alerte ou de proposition alternative pour refuser de mettre en œuvre une consigne qu'on leur demande d'appliquer, 62% répondent oui.

DC: pour les experts en sciences humaines et sociales, ce que fait l'Inria avec Transalgo est très important parce qu'on change de *régime probatoire*: jusqu'à présent les critiques d'algorithmes par des utilisateurs, par des institutions, ce sont des copies d'écran sous contrôle d'huissier. C'est insuffisant pour inquiéter ceux qui pourraient l'être ainsi que leurs avocats. On a donc besoin d'algorithmes pour critiquer les algorithmes. Développer un savoir, une ingénierie sur ces questions, c'est un moyen de les apprivoiser, de faire société avec, de savoir être vigilant tout en acceptant leur potentialité et le fait qu'il faut des solutions informatiques pour observer l'informatique.

2.2 Le pluralisme face à l'enfermement algorithmique

JF: le risque d'enfermement, c'est l'idée que la personnalisation des services que permettent les algorithmes serait opposée au pluralisme, les algorithmes augmentant la propension des individus à ne fréquenter que des objets, des personnes, des opinions, des cultures conformes à leurs propres goûts et à rejeter l'inconnu. Question à DC, en tant que sociologue, l'endogamie, le fait de rester entre soi n'est pas une nouveauté liée aux algorithmes, quel sont les éléments nouveaux dans ce risque d'enfermement lié aux algorithmes?

DC: le débat n'est pas nouveau, (je le dis pour BP) il y a eu ce débat pour l'arrivée des chaînes du câble : était évoquée l'idée que cela allait enfermer les gens dans des communautés, car ils ne regarderaient plus la grande messe du soir des grands journaux sur les chaînes principales. L'idée d'un enfermement n'est donc pas nouvelle. Tout d'abord, beaucoup d'algorithmes de recommandation, ceux qui fonctionnent efficacement, ont bien prélevé des traces du contexte de l'utilisateur, pour lui reconduire des choses qu'il a l'habitude de faire. De cette façon, se créent des boucles classiques de régularité où des comportements monotones informent des calculateurs pour faire des prévisions d'un futur qui ressemble terriblement à un passé calculé.

En fait, il y a des principes multiples qui dirigent aujourd'hui un algorithme et il est difficile de poser le débat en termes d'un grand principe dont on comprendrait la logique, que l'on pourrait discuter...

L'algorithme n'est rien de plus que le couple entre un artefact et des pratiques sociales. Le numérique a par ailleurs apporté une diversité informationnelle comme jamais connue dans toute l'Histoire de l'Humanité. Il est absurde de dire que Facebook enferme les gens.

Mais cela soulève des dangers : des gens curieux vont envoyer des signaux de curiosité et vont se voir incités en retour à la curiosité. En revanche, des gens donnant peu de traces de curiosité vont être dirigés vers moins de diversité. Si les traces données à l'algorithme sont variées et diversifiées, alors celui-ci enrichira les recommandations avec des nouveautés en périphérie et donc ouvrira l'utilisateur à des mondes divers. Des comptes FB avec des utilisateurs monotones, réguliers, qui se ressemblent produisent très peu de variété. Tandis que des comptes plus ouverts à une diversité relationnelle forte vont produire de la diversité.

Aujourd'hui dans le monde éditorial et de la culture, nous n'avons jamais publié autant de livres, cela ne veut pas dire que les livres lus ne sont pas très concentrés, peut-être qu'ils sont tous écrits de la même façon, avec le même type d'histoire ou de structure etc... La diversité implique une variété équilibrée, avec de nombreux acteurs/distributeurs, mais aussi une distance entre les produits et les informations. Un risque existe pour que se produise dans un certain contexte et pour un certain public, des pratiques sociales dans lesquelles l'algorithme ne sera pas un facteur d'enrichissement et de découverte, mais plutôt de reconduction du monde.

BP: Bien sûr, il y n'y a jamais eu autant de biens culturels, autant de séries, de films et de livres produits. Mais le phénomène de concentration sur un nombre de produits culturels similaires n'a jamais été aussi fort. Est-ce lié aux algorithmes? En tout cas c'est un constat qui n'est pas étranger au phénomène de contextualisation. Il y a bien un mécanisme de *winner takes it all*, dans les productions culturelles et informationnelles.

Comme DC, je ne crois pas que nous soyons enfermés, mais plutôt qu'apparaît une forme de délégation volontaire du choix à opérer, tandis que l'enfermement, c'est ne jamais être mis au contact de quelque chose d'autre, ce qui n'est pas le produit des algorithmes. Le choix laissé à l'outil ne nous enferme pas, mais reproduit un schéma de conduite passée.

En ce qui concerne l'information, on manque encore d'études. Mais un changement s'opère, qui est indirectement lié aux algorithmes. Par exemple, de récents travaux montrent que la réceptivité d'un article d'information sur Facebook est d'autant plus forte que cette information n'est pas envoyée par un média, mais au contraire partagée par l'un de vos amis. Une forme de « blanchiment » de l'information, que celle-ci soit légitime ou non au départ, conditionne de plus en plus la réception de l'information.

Aujourd'hui en journalisme, on vérifie moins des faits que des sources. La validation d'une source légitime l'information. Les algorithmes participent à la fragmentation.

AR: cette question de la bulle filtrante n'est pas propre aux algorithmes : nous sommes des êtres très prévisibles, aux comportements très réguliers, facilitant la possibilité de nous enfermer dans des bulles. Mais on ne nous enferme que si c'est rentable. Tout est une question de paramétrage des algorithmes. Ils peuvent aussi, au contraire nous exposer à des éléments ou à des informations que nous n'aurions jamais cherché à consulter.

Je crois, comme Musil, que la liberté se résume pour l'être humain à faire volontairement ce qu'il veut involontairement. Comment penser la liberté dans ce monde déterministe avec les algorithmes?

L'endroit où peut se manifester la liberté, c'est dans l'après coup, en étant capables de considérer ce qui nous a amenés à tel ou tel agissement, désir etc... C'est dans ces espaces hétérotopiques et à des moments hétérochroniques, où l'on réfléchit a posteriori sur ce qui nous a poussés à agir, que nous sommes des sujets de droit.

Cela pousse à élargir la vision dichotomique classique hommes/machines, avec les machines d'un côté, qui deviendraient autonomes, auxquelles on demande de rendre compte (trajectoire de l'aveu), et nous, êtres humains, qui nous n'aurions plus à rendre compte, la machine prenant en main ce qui nous fait agir, quels sont nos désirs etc... Détectés et prédits par les algorithmes. Tout ceci conduit à une obsession pour la loyauté, la transparence des algorithmes etc... Mais les trois lois de la robotique, à qui s'adressent-elles finalement? Aux êtres humains. L'IA étant conditionnée par et pour les êtres humains.

Il faut a posteriori, garder l'espace et le temps de raconter, de façon langagière, avec des mots, ce qui nous fait agir. Cette narrativité est essentielle, car elle permet aussi l'espace public, à travers des échanges à propos de la chose publique, non rabattue sur la seule concurrence des intérêts individuels égoïstes.

BP : en économie, les biens culturels étaient traditionnellement considérés comme des biens d'expérience, des biens dont on ne pouvait pas calculer l'utilité pour le consommateur avant qu'ils soient expérimentés. Les algorithmes ont bouleversé cela, expliquant peut-être pourquoi il y a un phénomène de surconcentration, avec la possibilité de calculer *ex ante* l'utilité de ces biens culturels.

NB : les algorithmes d'apprentissage sont des systèmes à réaction positive, qui perpétuent un passé. Se pose la question de savoir comment renforcer la pluralité pour limiter les biais d'apprentissages de certains algorithmes prédictifs.

Question : Le web étant un lieu de construction collective, au lieu d'utiliser des algorithmes pour contrôler des algorithmes, ne pourrait-on pas mettre en avant la place occupée par des collectifs qui, quotidiennement – Wikileaks, Panama Papers par exemple – font un pied de nez subversif aux GAFA, phénomène qui s'inscrit a contrario du sentiment d'avoir à faire à tes utilisateurs exclusivement seuls, face à leur écran et aux algorithmes? (D'un point de vue sociologique)

Question de Marc REES (NextImpact) : Pourquoi n'avons-nous pas de représentants des entreprises Google, Facebook (GAFA) ?

Question d'Alexei Grinbaum (CERNA) : Il y a probablement un quatrième terme à ajouter aux trois autres évoqués, loyauté, transparence et pluralité. C'est celui d'*explicabilité*. En effet, le vocabulaire propre à la machine ne correspond pas au sens du langage commun, à la sémantique humaine, et c'est constitutif d'un problème majeur dans la relation homme-machine, qu'il faudrait également traiter.

DC: dire qu'il faut des algorithmes pour critiquer les algorithmes, c'est dire qu'il faut un nouveau régime probatoire, applicable aux collectifs. L'idée que des initiatives se mettent en place pour contrôler les systèmes algorithmiques, les tromper, fausser les données et ce, de manière collective ne fait aucun doute et s'opère déjà naturellement dans la société. Ce nouveau régime probatoire est en train de se mettre en place.

Cependant, ce rapport tactique à la puissance de calcul, même dans des communautés, a besoin des supports de régulation.

Quant de l'explicabilité, même dans le *deep learning*, il y a la possibilité de revenir à de l'explicabilité. L'enjeu, c'est de faire de l'explication à destination du public (par exemple, pourquoi FB me recommande tel ami ?).

NB: concernant l'aspect collectif, dans Transalgo, il y a un comité de chercheurs, mais l'entourage se veut pluridisciplinaire, Maths, Info, mais aussi Sciences Humaines et Sociales, Juridique, Sociologique, Economique ainsi qu'avec des institutions CNnum, Cnil, associations de consommateurs, FING. Ces questions-là nécessitant des acteurs du plus large panel possible.

AR: l'apparition des collectifs est positive et effrayante, car symptomatique d'une désertion par le politique institutionnel des questions qui concernent la survivance de la vie. Toute la mise en place de ces collectifs reflète bien la faillite à laquelle nous mène la *gouvernementalité algorithmique*, système qui immunise le calculable, l'optimisation, la maximisation des profits etc... contre la vie elle-même, contre les processus d'altération, contre le non-optimisable, l'altération qui est au principe même de tout ce qui vit, comme le disait Canguilhem, et la normativité de la vie même. On a une sorte de normativité quantitative, qui se présente comme une immunisation du calculable contre la normativité vivante qui est précisément incalculable dans la mesure où s'il y a une norme propre au vivant c'est précisément parce que le vivant est infini, imprévisible, et c'est pourquoi il y a une normativité inhérente au vivant, toujours d'après Canguilhem.

Le problème de la *gouvernementalité algorithmique*, c'est qu'elle nous dispense de gouverner. Et cette fonction de gouverner, elle doit bien être prise en charge : cela débouche sur la constitution de ces collectifs, avec une efficacité certaine, rendue possible parce qu'il y a une différence essentielle de niveau, entre le niveau de sémiotique signifiante sur internet par lesquelles les gens s'expriment dans une langue compréhensible par tous et l'autre côté, le côté obscur de la machine, obscur dans le sens non-intelligible, qui est le côté des données décontextualisées, rendues insignifiantes et purement calculables, le côté machine qui fonctionne sur des signaux a-signifiants, au lieu de fonctionner sur base de sémiotique signifiante, donc de sens construit et gouvernable.

Isabelle Falque-Pierrotin : Merci aux participants et à l'assemblée, présentation du site éthique et numérique où l'on peut suivre et retrouver une vue d'ensemble des manifestations de tous les partenaires à ce débat public, manifestations qui seront de natures variées et différentes, afin de recueillir des discours experts, mais également plus profanes sur cette question de l'éthique et des algorithmes.