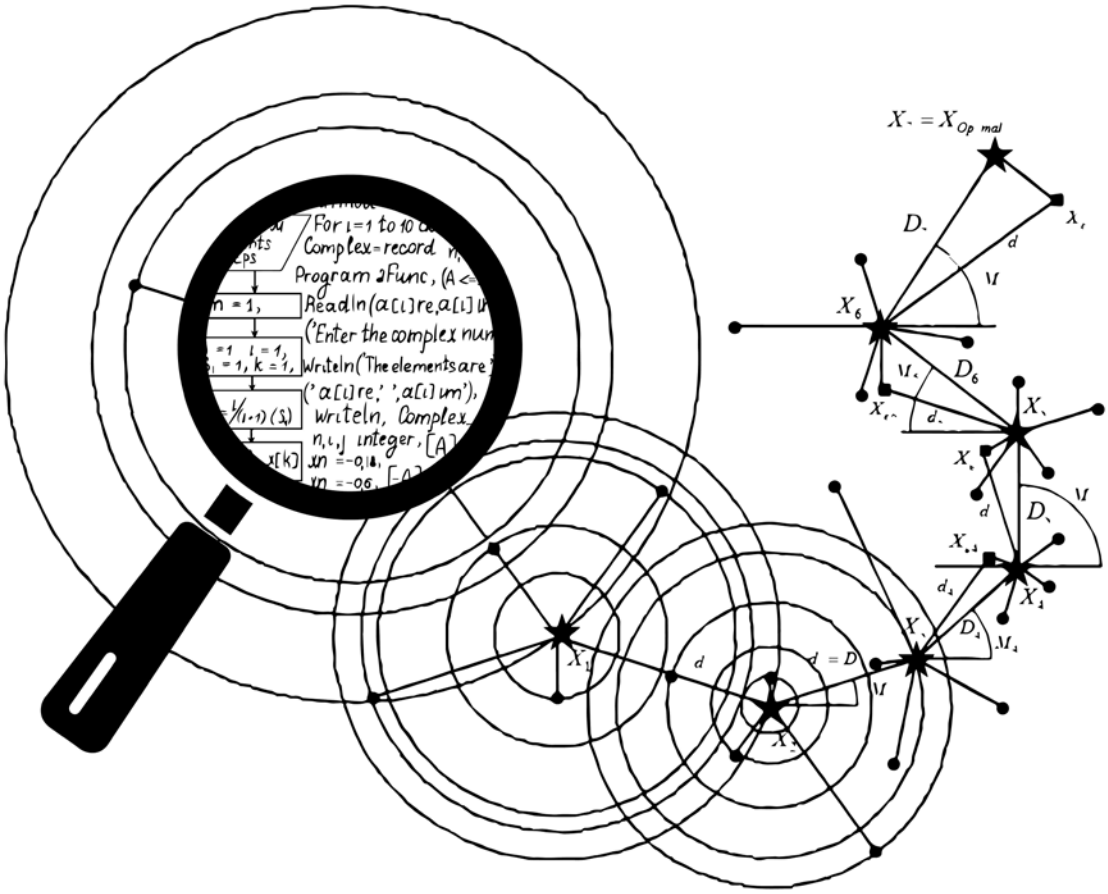


# COMMENT PERMETTRE À L'HOMME DE GARDER LA MAIN ?

## Les enjeux éthiques des algorithmes et de l'intelligence artificielle

SYNTHÈSE DU DÉBAT PUBLIC ANIMÉ PAR LA CNIL DANS LE CADRE DE LA MISSION DE RÉFLEXION ÉTHIQUE CONFIÉE PAR LA LOI POUR UNE RÉPUBLIQUE NUMÉRIQUE

**DÉCEMBRE 2017**

CNIL.
COMMISSION NATIONALE INFORMATIQUE & LIBERTÉS

et éthique numérique
les algorithmes en débat

# COMMENT PERMETTRE À L'HOMME DE GARDER LA MAIN ?

## Les enjeux éthiques des algorithmes et de l'intelligence artificielle

SYNTHÈSE DU DÉBAT PUBLIC ANIMÉ PAR LA CNIL DANS LE CADRE DE LA MISSION
DE RÉFLEXION ÉTHIQUE CONFIÉE PAR LA LOI POUR UNE RÉPUBLIQUE NUMÉRIQUE

**DÉCEMBRE 2017**

*Ce rapport a été rédigé par Victor Demiaux avec le concours de Yacine Si Abdallah.*

# PRÉFACE

L'intelligence artificielle est le grand mythe de notre temps. L'un annonce la destruction en masse de nos emplois, un autre l'émergence apocalyptique d'une conscience robotique hostile, un troisième la ruine d'une Europe écrasée par la concurrence. D'autres encore nourrissent plutôt le rêve d'un monde sur mesure, d'un nouvel Âge d'or d'où toute tâche ingrate ou répétitive serait bannie et déléguée à des machines ; un Eden où des outils infaillibles auraient éradiqué la maladie et le crime, voire le conflit politique, en un mot aboli le mal. Sous ses avatars tour à tour fascinants ou inquiétants, solaires ou chtoniens, l'intelligence artificielle dit sans doute plus de nos phantasmes et de nos angoisses que de ce que sera notre monde demain. À considérer l'attrait de ce type de discours eschatologiques en Europe, on en vient à penser que la technique cristallise aussi une puissance de projection dans l'avenir qui fait parfois défaut à nos imaginaires politiques.

Désamorcer ces présentations sensationnalistes des nouvelles technologies est une chose. Cela ne signifie pas pour autant que l'on ignore que l'irruption dans nos vies quotidiennes de ces assistants ou outils d'un nouveau type génère des bouleversements multiples et des défis nouveaux que nous devons relever. Préservation de l'autonomie de la décision humaine face à des machines parfois perçues comme infaillibles, détection de discriminations générées involontairement par des systèmes mouvants, sauvegarde de logiques collectives parfois érodées par la puissance de personnalisation du numérique, etc. : les enjeux ne manquent pas, aux implications déjà tangibles. Ils questionnent certains des grands pactes et des équilibres sur lesquels repose notre vie collective.

Établir de façon claire et lucide ces enjeux est le premier devoir de la puissance publique, la condition pour pouvoir proposer des réponses adaptées, pour intégrer l'innovation technologique à la construction d'une vision déterminée de notre avenir. C'était le sens de la création de la mission de réflexion sur les enjeux éthiques soulevés par les technologies numériques que la Loi pour une République numérique a confiée à la CNIL.

Comment appréhender aujourd'hui une telle mission ? Beaucoup se sont interrogés, voire ont questionné cette responsabilité nouvelle de la Commission. Comment exprimer l'éthique sur des sujets hautement complexes et évolutifs, à quel titre, selon quelles modalités ?

Réflexion sur les principes fondamentaux de conduite de la vie des hommes et des sociétés, définition d'un pacte social partagé sur un sujet complexe à un moment donné, l'éthique constitue un objet éminemment collectif, pluriel. Dans le domaine bien particulier des sciences de la vie et de la santé, la composition et la collégialité du travail du Comité Consultatif National d'Éthique répondent à cet impératif de pluralité.

Garante de principes éthiques fixés par le législateur il y a quarante ans, la CNIL a certes toute légitimité pour être le point focal de cette réflexion éthique, à l'heure où des possibilités techniques nouvelles soulèvent de nouveaux enjeux ou questionnent les équilibres antérieurs.

En revanche, il est apparu impensable qu'elle puisse se prévaloir d'un quelconque monopole sur la réflexion éthique du numérique. Sur un sujet aussi vaste et transversal, cette dernière ne saurait se concevoir en vase clos. Le numérique n'est pas un secteur, que l'on pourrait confier aux soins d'un comité d'éthique restreint à quelques membres, aussi compétents soient-ils. Il fallait innover.

C'est dans cet esprit que la CNIL a suscité une démarche collective, en animant avec l'aide de partenaires de multiples secteurs un débat public pendant plusieurs mois. L'éthique est à cet égard autant un processus d'élaboration que le résultat du processus lui-même. Nous avons ainsi fait le choix de partir des usages, des interrogations existantes et des pistes de solutions évoqués par les acteurs du débat. Plus de quarante manifestations organisées à Paris et en régions ont permis de recueillir les éléments qui ont alimenté le présent rapport et les recommandations dont il est porteur.

Un effort d'innovation était également nécessaire pour faire droit à la nécessité d'associer davantage le citoyen à l'élaboration de la réflexion publique sur un univers complexe qui modèle de plus en plus son existence et implique des choix de société fondamentaux. Un univers dont il doit être de plus en plus un co-acteur. La CNIL a ainsi organisé une journée de concertation citoyenne, à Montpellier, le 14 octobre dernier, qui a permis de joindre la voix d'une quarantaine de volontaires à la polyphonie du débat public.

Le premier bénéfice de cette démarche ouverte et décentralisée est d'avoir fait respirer le débat le plus largement possible et d'avoir participé à la montée en compétence de la société française vis-à-vis des questions soulevées par les algorithmes et par l'IA. Face à des systèmes socio-techniques de plus en plus complexes et compartimentés, face aux impacts parfois difficilement prévisibles d'artefacts en évolution constante, cloisonner le débat à quelques cercles d'initiés, c'est prendre le risque de susciter méfiance et défiance. Faire de l'ensemble de nos concitoyens des utilisateurs éclairés et critiques des technologies est au contraire un impératif tout à la fois éthique, démocratique et pragmatique. C'est aussi, pour la CNIL, prolonger l'œuvre d'accompagnement de la rencontre de la société française avec le numérique qu'elle accomplit depuis 40 ans.

À l'heure même où se définit la position française – et bientôt européenne – en matière d'intelligence artificielle, le rapport issu de ces mois de débat public contribue à poser les jalons d'un questionnement commun. Il propose un panorama des enjeux et formule un certain nombre de principes et de recommandations.

Celles-ci ont un objectif commun : permettre à la personne humaine de ne pas « perdre la main ». À l'heure de la dématérialisation généralisée, ceci paraitra peut-être décalé. Il nous semble au contraire que c'est là que réside notre défi collectif majeur. Faire en sorte que ces nouveaux outils soient à la main humaine, à son service, dans un rapport de transparence et de responsabilité.

Puissent ces réflexions alimenter celles en cours au sein des pouvoirs publics, dont celle de la mission Villani, mais aussi des différentes composantes de la société civile. Puissent-elles ainsi participer à l'élaboration d'un modèle français de gouvernance éthique de l'intelligence artificielle.

**Isabelle Falque-Pierrotin**
Présidente de la CNIL

# SOMMAIRE

# RÉSUMÉ

Ce rapport est le résultat d'un débat public animé par la CNIL. Entre janvier et octobre 2017, 60 partenaires (associations, entreprises, administrations, syndicats, etc.) ont organisé 45 manifestations dans toute la France. Il s'agissait d'identifier les sujets de préoccupations éthiques soulevés par les algorithmes et l'intelligence artificielle, ainsi que les pistes de solutions possibles.

La première partie du rapport apporte une définition pragmatique des algorithmes et de l'intelligence artificielle tout en présentant leurs principaux usages et notamment ceux d'entre eux qui retiennent aujourd'hui le plus l'attention publique. Classiquement, l'algorithme se définit ainsi comme une suite finie et non ambigüe d'instructions permettant d'aboutir à un résultat à partir de données fournies en entrée. Cette définition rend compte des multiples applications numériques qui, exécutant des programmes traduisant eux-mêmes en langage informatique un algorithme, remplissent des fonctions aussi diverses que fournir des résultats sur un moteur de recherche, proposer un diagnostic médical, conduire une voiture d'un point à un autre, détecter des suspects de fraude parmi les allocataires de prestations sociales, etc. L'intelligence artificielle désigne principalement dans le débat public contemporain une nouvelle classe d'algorithmes, paramétrés à partir de techniques dites d'apprentissage : les instructions à exécuter ne sont plus programmées explicitement par un développeur humain, elles sont en fait générées par la machine elle-même, qui « apprend » à partir des données qui lui sont fournies. Ces algorithmes d'apprentissage peuvent accomplir des tâches dont sont incapables les algorithmes classiques (reconnaître un objet donné sur de très vastes corpus d'images, par exemple). En revanche, leur logique sous-jacente reste incompréhensible et opaque y compris à ceux qui les construisent.

**Le débat public a permis d'identifier 6 grandes problématiques éthiques :**

- Le perfectionnement et l'autonomie croissante des artefacts techniques permettent des formes de délégations de tâches, de raisonnements et de décisions de plus en complexes et critiques à des machines. Dans ces conditions, à côté de l'augmentation de sa puissance d'agir permise par la technique, n'est-ce pas aussi son autonomie, son libre arbitre, qui peut se trouver érodé ? Le prestige et la confiance accordés à des machines jugées souvent infaillibles et « neutres » ne risquent-ils pas de générer la tentation de se décharger sur les machines de la fatigue d'exercer des responsabilités, de juger, de prendre des décisions ? Comment appréhender les formes de dilution de la responsabilité que sont susceptibles de susciter les systèmes algorithmiques, complexes et très segmentés ?

- Les algorithmes et l'intelligence artificielle peuvent susciter des biais, des discriminations, voire des formes d'exclusion. Ces phénomènes peuvent être volontaires. Mais le réel enjeu, à l'heure du développement des algorithmes d'apprentissage, est leur développement à l'insu même de l'homme. Comment y faire face ?

- L'écosystème numérique tel qu'il s'est construit avec le Web, mais également plus anciennement les techniques actuarielles, ont fortement exploité les potentialités des algorithmes en termes de personnalisation. Le profilage et la segmentation de plus en plus fine rendent bien des services à l'individu. Mais cette logique de personnalisation est également susceptible d'affecter – outre les individus – des logiques collectives essentielles à la vie de nos sociétés (pluralisme démocratique et culturel, mutualisation du risque).

- L'intelligence artificielle, dans la mesure où elle repose sur des techniques d'apprentissage, nécessite d'énormes quantités de données. Or, la législation promeut une logique de minimisation de la collecte et de la conservation de données personnelles, conforme à une conscience aigüe des risques impliqués pour les libertés individuelles et publiques de la constitution de grands fichiers. Les promesses de l'IA justifient-elles une révision de l'équilibre construit par le législateur ?

- Le choix du type de données alimentant un modèle algorithmique, leur quantité suffisante ou insuffisante, l'existence de biais dans les jeux de données servant à entraîner les algorithmes d'apprentissage constituent un enjeu majeur. S'y cristallise le besoin d'établir une attitude critique et de ne pas nourrir une confiance excessive dans la machine.

- L'autonomie croissante des machines ainsi que l'émergence de formes d'hybridation entre humains et machines (hybridation au plan d'une action assistée par des recommandations algorithmiques, mais aussi prochainement au plan physique) questionnent l'idée d'une spécificité humaine irréductible. Faut-il et est-il possible de parler au sens propre d' « éthique des algorithmes » ? Comment appréhender cette nouvelle classe d'objets que sont les robots humanoïdes, objets mais susceptibles de susciter chez l'homme des formes d'affects et d'attachement ?

La troisième partie du rapport envisage les réponses possibles formulées à l'occasion du débat public.
Elle aborde d'abord les principes susceptibles de construire une intelligence artificielle au service de l'homme. Deux principes nouveaux apparaissent comme fondateurs.

Le premier, substantiel, est le ***principe de loyauté,*** dans une version approfondie par rapport à celle initialement formulée par le Conseil d'Etat sur les plateformes. Cette version intègre en effet une dimension collective de la loyauté, celle-ci visant à ce que l'outil algorithmique ne puisse trahir sa communauté d'appartenance (consumériste ou citoyenne), qu'il traite ou non des données personnelles.

Le second, d'ordre plus méthodique, est un **principe de vigilance/réflexivité.** Il vise à répondre dans le temps au défi constitué par le caractère instable et imprévisible des algorithmes d'apprentissage. Il constitue aussi une réponse aux formes d'indifférence, de négligence et de dilution de responsabilité que peut générer le caractère très compartimenté et segmenté des systèmes algorithmiques. Il a enfin pour but de prendre en compte et de contrebalancer la forme de biais cognitif conduisant l'esprit humain à accorder une confiance excessive aux décrets des algorithmes. Il s'agit d'organiser, par des procédures et mesures concrètes, une forme de questionnement régulier, méthodique, délibératif et fécond à l'égard de ces objets techniques de la part de tous les acteurs de la chaine algorithmique, depuis le concepteur, jusqu'à l'utilisateur final, en passant par ceux qui entraînent les algorithmes.

Ces deux principes apparaissent comme fondateurs de la régulation de ces outils et assistants complexes que sont les algorithmes et l'IA. Ils en permettent l'utilisation et le développement tout en intégrant leur mise sous contrôle par la communauté.

Ils sont complétés par une ingénierie spécifique et nouvelle articulée sur deux points : l'un visant à repenser l'obligation d'intervention humaine dans la prise de décision algorithmique (article 10 de la loi Informatique et libertés) ; l'autre à organiser l'intelligibilité et la responsabilité des systèmes algorithmiques.

Ces principes font ensuite l'objet d'une déclinaison opérationnelle sous la forme de **6 recommandations** adressées tant aux pouvoirs publics qu'aux diverses composantes de la société civile (grand public, entreprises, associations, etc.) :

- Former à l'éthique tous les maillons de la « chaîne algorithmique (concepteurs, professionnels, citoyens) ;
- Rendre les systèmes algorithmiques compréhensibles en renforçant les droits existants et en organisant la médiation avec les utilisateurs ;
- Travailler le *design* des systèmes algorithmiques au service de la liberté humaine ;
- Constituer une plateforme nationale d'audit des algorithmes ;
- Encourager la recherche sur l'IA éthique et lancer une grande cause nationale participative autour d'un projet de recherche d'intérêt général ;
- Renforcer la fonction éthique au sein des entreprises.

# Une démarche innovante au service de l'élaboration d'une réflexion éthique collective et pluraliste

## Un débat public national sur les enjeux éthiques des algorithmes et de l'intelligence artificielle

La loi pour une République numérique de 2016 a confié à la CNIL la mission de conduire une réflexion sur les enjeux éthiques et les questions de société soulevés par l'évolution des technologies numériques.

La CNIL a choisi de faire porter en 2017 cette réflexion sur les algorithmes à l'heure de l'intelligence artificielle. En effet, ceux-ci occupent dans nos vies une place croissante, bien qu'invisible. Résultats de requêtes sur un moteur de recherche, ordres financiers passés par des robots sur les marchés, diagnostics médicaux automatiques, affectation des étudiants à l'Université : dans tous ces domaines, des algorithmes sont à l'œuvre. En 2016, le sujet des algorithmes s'était d'ailleurs invité de manière inédite dans le débat public et a suscité une forte attention médiatique (questions sur l'algorithme du logiciel Admission Post-Bac, recours à l'intelligence artificielle dans la stratégie électorale du candidat Trump, rôle des réseaux sociaux dans la diffusion des « fake news »).

> La réflexion éthique porte sur des choix de société décisifs. Elle ne saurait se construire indépendamment d'une prise en compte de cette dimension pluraliste et collective

La réflexion éthique porte sur des choix de société décisifs. Elle ne saurait se construire indépendamment d'une prise en compte de cette dimension pluraliste et collective. Ceci est d'autant plus vrai quand il s'agit d'un objet aussi transversal à toutes les dimensions de notre vie individuelle et sociale que les algorithmes. Il ne serait guère envisageable de rassembler en un unique comité l'ensemble des compétences et des regards nécessaire à l'examen des enjeux soulevés par les algorithmes dans des secteurs aussi divers que la santé, l'éducation, le marketing, la culture, la sécurité, etc.

Plutôt que de conduire directement sur ces sujets une réflexion centralisée, la CNIL a donc fait le choix de se positionner, d'une façon originale, en tant qu'animatrice d'un débat public national ouvert et décentralisé. À l'occasion d'un événement de lancement organisé le 23 janvier 2017, elle a ainsi appelé tous les acteurs et organismes – institutions publiques, société civile, entreprises – qui le souhaitaient à organiser un débat ou une manifestation sur le sujet, dont ils lui feraient ensuite parvenir la restitution. L'objectif a donc été de s'adresser aux acteurs de terrain pour recueillir auprès d'eux les sujets éthiques identifiés comme tels à ce jour ainsi que les pistes de solutions évoquées par les uns et par les autres.

Soixante partenaires ont souhaité répondre à l'appel lancé par la CNIL. De natures très diverses, ces acteurs relevaient de secteurs très différents. Citons, à titre d'exemples, la Ligue de l'Enseignement dans l'éducation, la Fédération Française de l'Assurance (FFA), le Ministère de la Culture (DGMIC), l'association Open Law, ou encore la CFE-CFC et FO Cadres (ressources humaines), etc. Ces 60 partenaires ont organisé 45 manifestations entre

mars et octobre 2017 dans plusieurs villes de France (mais également à l'étranger grâce à la participation de la « Future Society at Harvard Kennedy School »), auxquelles ont participé environ 3000 personnes. La CNIL a assuré la coordination et la mise en cohérence de l'ensemble.

Les manifestations organisées dans le cadre du débat public ont aussi constitué l'occasion **de faire vivre dans la société française la réflexion sur des enjeux dont la prise de conscience par l'ensemble de nos contemporains, et pas seulement par les experts, est un enjeu civique et démocratique capital.**

## Une concertation citoyenne : Montpellier, 14 octobre 2017

Les questions posées par les algorithmes et l'intelligence artificielle renvoient à des choix de société et concernent tous les citoyens. L'organisation d'une concertation a donc eu pour objectif de recueillir le point de vue de simples citoyens. Il s'agissait de compléter les réflexions émises à l'occasion de diverses manifestations ayant principalement donné la parole à des experts de différents secteurs.

Une journée de concertation a ainsi été organisée le 14 octobre 2017, avec le soutien de la Ville de Montpellier et de Montpellier Méditerranée Métropole. Un appel à candidature a permis de recruter un panel diversifié de 37 citoyens.

Le format retenu visait à **favoriser l'échange d'idées et la construction d'un avis collectif.** La technique d'animation a permis successivement aux participants de :

- Comprendre ce que sont les algorithmes et l'intelligence artificielle ;
- Analyser collectivement quatre études de cas (médecine et santé / ressources humaines / personnalisation et enfermement algorithmique / éducation et transparence) pour identifier les opportunités et les risques liés à l'usage des algorithmes ;
- Formuler des recommandations pour assurer le déploiement dans un cadre éthique des algorithmes et de l'IA, le degré de consensus de celles-ci ayant ensuite été évalué.

Les résultats et enseignements sont présentés dans les encadrés « Le regard du citoyen ».

## La composition du rapport

Les manifestations organisées par les partenaires, ainsi que la concertation citoyenne, ont fait l'objet de restitutions recueillies par la CNIL. Les réflexions émises par des acteurs pluriels (syndicats, associations, entreprises, chercheurs, citoyens, etc.) dans des secteurs très divers (de l'assurance à l'éducation, en passant par la justice et la santé) ont ainsi alimenté le présent rapport, qui constitue un panorama général des questions éthiques soulevées par les algorithmes et l'intelligence artificielle dans leurs applications actuelles et dans leurs promesses à relativement court terme.

Animatrice du débat public, la CNIL en est aussi la restitutrice. À cet égard, elle a assumé la composition du rapport, ce qui implique inévitablement certains choix. La ligne de conduite adoptée a consisté à rendre loyalement et pleinement compte de la pluralité des points de vue exprimés. C'est aussi ce qui explique que les recommandations formulées à la fin du rapport entendent moins clore le débat que laisser ouvertes un certain nombre d'options possibles (dimension incitative ou obligatoire des mesures proposées, par exemple) qui devraient faire l'objet d'arbitrages ultérieurs. Il s'agit donc d'éclairer la décision publique et non de s'y substituer.

La CNIL s'est également appuyée pour la rédaction du rapport sur un travail de recherche documentaire, souvent initié sur la recommandation de tel ou tel partenaire. Les articles ou ouvrages utilisés ont été mentionnés en notes de bas de page. On pourra également se reporter aux pages du site de la CNIL consacrées au débat éthique pour retrouver quelques éléments de bibliographie sommaire[1]. Enfin, ont été exploités les résultats d'un certain nombre de travaux déjà conduits par diverses institutions en France et à l'étranger (entre autres, l'OPECST, la CERNA, le CNNUM, le Conseil d'Etat, la CGE, la Maison Blanche, France IA, INRIA, AI Now).

> Les questions posées par les algorithmes et l'intelligence artificielle renvoient à des choix de société et concernent tous les citoyens

1 https://www.cnil.fr/fr/ethique-et-numerique-les-algorithmes-en-debat-1

# LES DATES CLÉS

**7 OCTOBRE 2016**

La CNIL obtient pour mission par la loi « République Numérique » de conduire une réflexion sur les enjeux éthiques et de société soulevés par les nouvelles technologies

**23 JANVIER 2017**

La CNIL annonce pour 2017 le thème des algorithmes et de l'intelligence artificielle et organise des tables-rondes de lancement réunissant des experts de ces sujets

**FIN MARS 2017**

Les premiers événements sont organisés par les partenaires du débat public

**DÉBUT OCTOBRE 2017**

**45 événements** se sont tenus, à l'initiative des **60 partenaires** du débat public

**14 OCTOBRE 2017**

La CNIL organise une concertation citoyenne à Montpellier réunissant près de **40 citoyens**

**15 DÉCEMBRE 2017**

La CNIL présente le rapport « **Comment permettre à l'Homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle** », synthèse du débat public

# LES CHIFFRES CLÉS

**45** ÉVÉNEMENTS

**60** PARTENAIRES

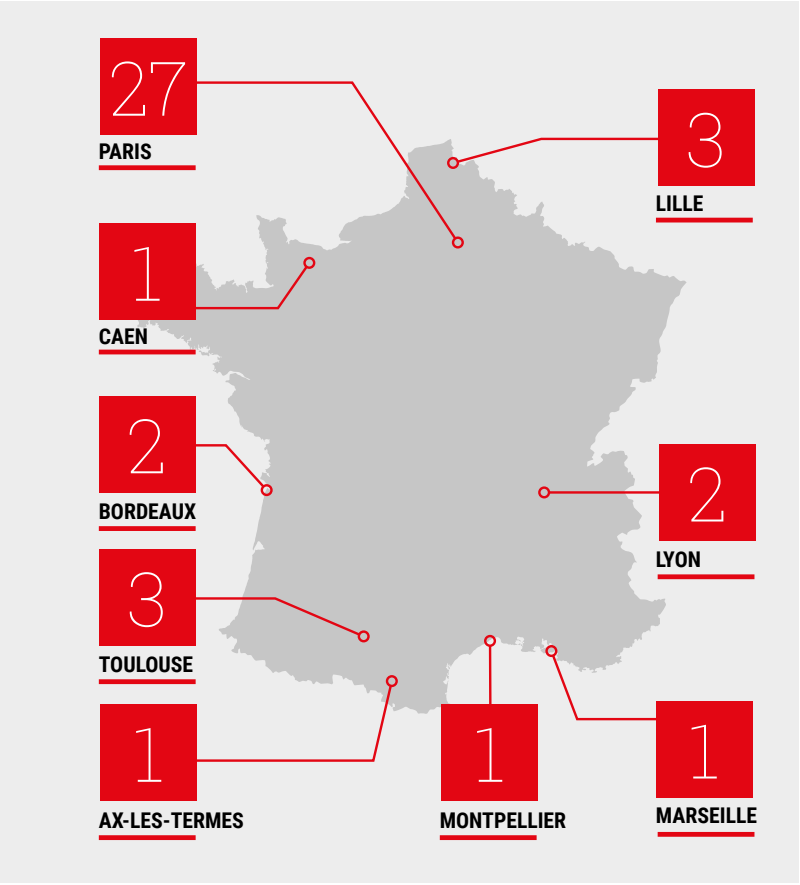
**1** JOURNÉE DE CONCERTATION CITOYENNE

ENVIRON **3 000** PERSONNES PRÉSENTES LORS DES MANIFESTATIONS

**27** À PARIS

**14** EN PROVINCE

**4** OUTRE ATLANTIQUE



# Algorithmes et Intelligence artificielle aujourd'hui

# Algorithmes et IA aujourd'hui

## Un effort de définition nécessaire à la qualité du débat public

Algorithmes et intelligence artificielle sont à la mode. Ces mots sont aujourd'hui partout, non sans confusion parfois. Les définitions et les exemples qui en sont donnés dans le débat public aujourd'hui sont souvent imprécis. Ils sont parfois même contradictoires. Cette situation s'explique par le caractère très technique de sujets qui se sont trouvés rapidement mis en circulation et en débat dans un espace public dépassant largement les cercles d'experts et de spécialistes auxquels ils se sont longtemps trouvés cantonnés.

De là, pour peu que l'on y prête attention, une extrême imprécision dans les termes employés. **Quoi de commun entre l'austère notion d' « intelligence artificielle » définie dans les milieux de la cybernétique dans les années 1950 et sa représentation populaire diffusée notamment par le cinéma hollywoodien ?** Qui prête d'ailleurs attention au fait qu' « intelligence » n'a pas la même signification en français et en anglais, langue dans laquelle a été créé le vocable *« artificial intelligence »* ? Comment comprendre que l'on dise ici que les algorithmes sont nouveaux et que d'autres voix nous assurent que l'homme y a recours depuis plusieurs milliers d'années ?

Outre les réalités et les projets techniques qu'ils entendent désigner, les algorithmes et l'intelligence artificielle en sont venus à constituer de nouvelles mythologies de notre temps, dont la simple évocation suffit à connoter la modernité et l'innovation numériques. Rien d'étonnant dès lors à ce que ces termes soient apposés de manière souvent rapide et peu justifiée à des réalités ou à des entreprises soucieuses de se forger une image flatteuse et futuriste : présenter son activité comme relevant du domaine de l'IA est aujourd'hui pour de nombreux acteurs un enjeu d'image, comparable à celui représenté depuis quelques années par l'invocation d'un « Big data » dont les spécialistes soulignent pourtant souvent qu'il demeure une réalité aux dimensions encore modestes. En tout état de cause, la réalité des promesses de l'IA est aujourd'hui un sujet de controverses plus ou moins explicites entre chercheurs en intelligence artificielle, entrepreneurs et prescripteurs d'opinions divers dans le domaine des technologies.

Comme on le rappellera par la suite, un autre type de confusion semble parfois entretenu par des acteurs dont l'activité est généralement reconnue comme relevant du domaine de l'intelligence artificielle. Ces derniers majoreraient résolument et exagérément non tant les promesses que les risques d'une intelligence artificielle qui parviendrait à s'autonomiser totalement de son créateur au point de mettre en danger l'humanité. Les voix les plus compétentes s'élèvent pour battre en brèche de telles prévisions, assimilées au mieux à des fantasmes, voire à des mensonges. Ceux-ci auraient pour fonction de détourner l'attention publique des

> Fonder une discussion publique saine et constructive sur les sujets des algorithmes et de l'intelligence artificielle nécessite absolument de préciser le rapport entre algorithmes et intelligence artificielle

problèmes certes plus prosaïques mais plus pressants soulevés par le déploiement de l'intelligence artificielle, en matière de lutte contre les discriminations ou de protection des données personnelles par exemple.

Disons-le d'emblée : toute définition en ces matières pourra être sujette à caution selon les différents points de vue. Dans la perspective du présent rapport, l'essentiel est de parvenir à une base de discussion minimale et opératoire qui permette de tracer pragmatiquement le périmètre des algorithmes et de l'intelligence artificielle sources de questions éthiques et de société cruciales. Autrement dit, il s'agit de proposer une définition aussi rigoureuse que possible mais prenant en compte la perception commune de ce en quoi algorithmes et IA constituent aujourd'hui des enjeux dignes d'attention.

**ENQUÊTE**

### Algorithmes et IA : un objet mal connu des Français*

Les algorithmes sont présents dans l'esprit des Français mais de façon assez confuse. Si **83% des Français ont déjà entendu parler des algorithmes,** ils sont **plus de la moitié à ne pas savoir précisément de quoi il s'agit** (52%). Leur présence est déjà appréhendée comme massive dans la vie de tous les jours par 80% des Français qui considèrent, à 65%, que cette dynamique va encore s'accentuer dans les années qui viennent.

**83 %**
**des Français ont déjà entendu parler des algorithmes**

* Sondage mené par l'IFOP pour la CNIL en janvier 2017 (auprès d'un échantillon de 1001 personnes, représentatif de la population française âgée de 18 ans et plus) sur le niveau de notoriété des algorithmes au sein de la population française.

# Les algorithmes : une réalité ancienne au cœur de l'informatique

Au sens strict, un algorithme est la description d'une suite finie et non ambigüe d'étapes (ou d'instructions) permettant d'obtenir un résultat à partir d'éléments fournis en entrée. Par exemple, une recette de cuisine est un algorithme, permettant d'obtenir un plat à partir de ses ingrédients[2]. L'existence d'algorithmes utilisés pour résoudre des équations est d'ailleurs attestée très anciennement, dès le IIIe millénaire en Mésopotamie babylonienne.

Dans le monde de plus en plus numérique dans lequel nous vivons, **les algorithmes informatiques permettent de combiner des informations les plus diverses pour produire une grande variété de résultats : simuler l'évolution de la propagation de la grippe en hiver, recommander des livres à des clients sur la base des choix déjà effectués par d'autres clients, comparer des images numériques de visages ou d'empreintes digitales, piloter de façon autonome des automobiles ou des sondes spatiales, etc.**

Pour qu'un algorithme puisse être mis en œuvre par un ordinateur, il faut qu'il soit exprimé dans un langage informatique, transcrit en un programme (une sorte de texte composé de commandes écrites, également appelé « code source »). Ce programme peut alors être exécuté dans un logiciel ou compilé sous la forme d'une application. Un logiciel a recours en général à de nombreux algorithmes : pour la saisie des données, le calcul du résultat, leur affichage, la communication avec d'autres logiciels, etc.

2 Voir par exemple : http://www.cnrtl.fr/definition/algorithme

# Des algorithmes à l'intelligence artificielle

**Peu de notions font aujourd'hui l'objet d'un usage plus mouvant que celle d' « intelligence artificielle » (IA).** Le choix a été fait dans ce rapport de se concentrer pragmatiquement sur les usages d'ores et déjà effectifs de l'intelligence artificielle et, plus précisément, sur ceux qui ont fait l'objet des plus rapides développement au cours des dernières années, en lien avec les progrès du *machine learning* (ou apprentissage automatique).

De façon large, l'intelligence artificielle peut être définie comme « la science qui consiste à faire faire aux machines ce que l'homme ferait moyennant une certaine intelligence » (Marvin Minsky). Si c'est en 1956, lors de la conférence de Darmouth que naît formellement la notion d'intelligence artificielle dans le milieu de la cybernétique, on peut considérer comme point de départ l'article publié en 1950 par Alan Turing *(Computing Machinery and Intelligence)* où celui-ci pose la question de savoir si les machines peuvent penser. Les chercheurs de cette discipline naissante ambitionnent de doter des ordinateurs d'une intelligence généraliste comparable à celle de l'homme, et non pas limitée à certains domaines ou à certaines tâches.

L'histoire de l'intelligence artificielle depuis les années 1950 n'a pas été celle d'un progrès continu. En premier lieu, les chercheurs se sont vus contraints de délaisser l'objectif visant à mettre au point une IA généraliste (ou « IA forte ») pour se concentrer sur des tâches plus spécifiques, sur la résolution de problèmes tels que la reconnaissance d'images, la compréhension du langage naturel ou la pratique de jeux (jeu de dames, échecs, jeu de go, par exemple). On parle dès lors d' « IA faible », car spécialisée. Même si l'on s'en tient au domaine de l'IA faible, l'histoire de ce champ de recherche et de ses applications est marquée par des discontinuités. À une période d'optimisme dans les années 1980 a succédé à partir des années 1990 un « hiver de l'IA » : les progrès se sont heurtés à une insuffisance tant de la puissance de calcul que des données disponibles, notamment.

**Ces dernières années ont au contraire été marquées par une série de succès spectaculaires qui ont remis au goût du jour les promesses de l'IA.** La victoire d'Alpha Go (Google) contre le champion du monde de jeu de Go, Lee Sedol, en mars 2016, a constitué symboliquement le plus notable de ces événements. Contrairement au jeu d'échecs, le go, du fait de la multiplicité innombrable des combinaisons qu'il permet, ne se prête pas à la mémorisation d'un grand nombre de parties que la machine pourrait se contenter de reproduire.

**La victoire d'Alpha Go illustre le fait que les développements récents de l'IA sont notamment liés au perfectionnement de la technique du *machine learning* (apprentissage automatique),** qui en constitue l'une des branches. Alors que le programmeur doit traditionnellement décomposer en de multiples instructions la tâche qu'il s'agit d'automatiser de façon à en expliciter toutes les étapes, l'apprentissage automatique consiste à alimenter la machine avec des exemples de la tâche que l'on se propose de lui faire accomplir. L'homme *entraîne* ainsi le système en lui fournissant des données à partir desquelles celui-ci va *apprendre* et déterminer lui-même les opérations à effectuer pour accomplir la tâche en question. Cette technique permet de réaliser des tâches hautement plus complexes qu'un algorithme classique. Andrew Ng, de l'Université Stanford, définit ainsi le *machine learning* comme « la science permettant de faire agir les ordinateurs sans qu'ils aient à être explicitement programmés ». Cela recouvre la conception, l'analyse, le développement et la mise en œuvre de méthodes permettant à une machine d'évoluer par un processus systématique, et de remplir des tâches difficiles. L'intelligence artificielle qui repose sur le *machine learning* concerne donc des algorithmes dont la particularité est d'être conçus de sorte que leur comportement évolue dans le temps, en fonction des données qui leur sont fournies.

L'apprentissage profond *(Deep learning)* est le socle des avancées récentes de l'apprentissage automatique, dont il est l'une des branches[3]. **On distingue apprentissage automatique supervisé[4]** (des données d'entrées qualifiées par des humains sont fournies à l'algorithme qui définit donc des règles à partir d'exemples qui sont autant de cas validés) **et non supervisé[5]** (les données sont fournies brutes à l'algorithme qui élabore sa propre classification et est libre d'évoluer vers n'importe quel état final lorsqu'un motif ou un élément lui est présenté). L'apprentissage supervisé nécessite que des instructeurs apprennent à la machine les résultats qu'elle doit fournir, qu'ils l'« entraînent ». Les personnes entraînant l'algorithme remplissent en fait souvent une multitude de tâches très simples. Des plateformes telles que le « Mechanical Turk » d'Amazon sont les lieux où se recrutent ces milliers de « micro-tâcherons » (Antonio Casilli) qui, par exemple, étiquettent les immenses

3 Il s'agit d'un ensemble de méthodes d'apprentissage automatique tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires. Sa logique étant inspirée du fonctionnement des neurones, on parle souvent de « réseaux neuronaux ».

4 Un algorithme de *scoring* de crédit utilisera cette technique : on fournit l'ensemble des caractéristiques connues des clients et de leur emprunt en indiquant ceux qui n'ont pas remboursé leur crédit, et l'algorithme sera capable de fournir un score de risque de non remboursement pour les futurs clients.

5 Un algorithme de détection des typologies de fraudes utilisera cette technique : on fournit à l'algorithme toutes les données relatives à des fraudes avérées, et l'algorithme sera capable de dégager des similitudes entre ces fraudes, et de dégager des typologies de fraudes. L'apprentissage non supervisé peut aussi servir à identifier, sur la bande sonore d'une émission de radio, les séquences de parole de différents locuteurs.

FOCUS

## L'exemple de la reconnaissance d'images

La reconnaissance d'images permet de prendre la mesure de ce qui distingue algorithmes classiques et algorithmes de *machine learning* (que le vocabulaire courant confond aujourd'hui généralement avec l'IA). Imaginons que l'on ait pour objectif de faire reconnaître les tigres à une machine. Si l'on se proposait d'y parvenir au moyen d'un algorithme classique, il faudrait imaginer pouvoir décrire explicitement en langage de programmation la totalité des opérations intellectuelles que nous réalisons lorsque nous identifions que nous avons à faire à un tigre et non pas, par exemple, à tout autre animal, voire à un lion ou à un chat. Si distinguer un tigre d'un chat ne pose aucun problème même à un jeune enfant, en décomposer et expliciter l'ensemble des étapes nécessaires à reconnaitre un tigre (autrement dit, en donner l'*algorithme*) s'avère être une tâche, sinon impossible du moins d'une ampleur rédhibitoire. C'est ici qu'intervient la technique du machine learning. Il s'agit de fournir à la machine des exemples en grande quantité, en l'occurrence de très nombreuses photographies de tigres, ainsi que des photographies d'autres animaux. À partir de ce jeu de données, la machine apprend à reconnaître des tigres. Elle élabore elle-même, par la confrontation des milliers de photographies qui lui sont fournies, les critères sur lesquels elle s'appuiera pour reconnaître des tigres dans des photographies qui lui seront ultérieurement soumises.

Il s'agit ici d' « apprentissage supervisé » : c'est bien l'homme qui fournit à la machine des milliers de photographies qu'il a préalablement identifiées comme représentant des tigres ainsi que d'autres explicitement identifiées comme ne représentant *pas* des tigres.

quantités de photographies utilisées pour entraîner un logiciel de reconnaissance d'images. Le système de captcha de Google « recaptcha » est un autre exemple d'utilisation à grande échelle d'humains pour entrainer des machines. Ces algorithmes d'apprentissage sont utilisés dans un nombre croissant de domaines, allant de la prédiction du trafic routier à l'analyse d'images médicales.

On comprend à travers l'exemple de la reconnaissance d'images (voir encadré) en quoi l'intelligence artificielle ouvre la voie à l'automatisation de tâches incomparablement plus complexes que l'algorithmique classique. L'IA, contrairement aux algorithmes déterministes construit elle-même à partir des données qui lui sont fournies les modèles qu'elle va appliquer pour appréhender les réalités qui lui sont soumises. Ainsi s'explique qu'elle s'avère aujourd'hui particulièrement prometteuse dans des secteurs produisant des quantités énormes de données, telles que la météorologie.

Les exemples d'utilisation de l'intelligence artificielle sont d'ores et déjà nombreux, bien au-delà du seul domaine de la reconnaissance de formes. Ainsi, la classification du spam parmi les messages reçus sur Gmail constitue une application caractéristique, dans sa simplicité même, de l'IA.

## LE SAVIEZ-VOUS ?

*Une entreprise comme Airbus mobilise aujourd'hui concrètement l'intelligence artificielle à des fins de reconnaissance de forme. Apprendre à un système à reconnaître sur une photographie aérienne d'une zone maritime les différents navires présents peut servir, par exemple, à confronter l'emplacement des embarcations ainsi repérées aux signaux émis par les balises et à identifier des navires en perdition ou qui cherchent à se soustraire à la surveillance maritime. L'intérêt réside dans la rapidité d'une opération qui, si elle n'est pas automatisée, réclame un temps et des moyens considérables. Depuis quelques années, les progrès de ces techniques sont tels que la machine surpasse désormais l'humain pour la fiabilité de l'identification de navires parfois difficilement distinguables de nuages.*

Le signalement par les usagers de messages considérés comme indésirables permet à Google de constituer une base conséquente et constamment alimentée à partir de laquelle le système peut apprendre à déterminer les caractéristiques des spams qui vont ensuite lui permettre de proposer de lui-même quels messages filtrer. Toujours chez Google, l'intelligence artificielle est à l'œuvre dans le service de traduction automatique. L'entreprise explique également avoir eu recours au *machine learning* pour analyser le fonctionnement du système de refroidissement de ses *data centers*. L'automatisation de cette fonction d'analyse aurait ainsi permis de réduire de 40 % l'énergie nécessaire au refroidissement de ces installations.

L'utilisation industrielle de l'IA n'est pas nouvelle : elle s'est notamment développée dans les années 1980, quand les « systèmes experts » ont permis d'optimiser l'opération de vidange des cuves des centrales nucléaires, automatisant les calculs et renforçant du même coup leur fiabilité en permettant de substantielles économies liées à la réduction de la durée d'immobilisation des installations à des fins de maintenance.

Les robots conversationnels (chat bots) et assistants vocaux (comme Siri, Google Assistant ou Alexa) constituent un autre pan en rapide développement de l'intelligence artificielle : ils peuvent par exemple fournir des informations et répondre à des questions standardisées.

À la lumière de ces applications, on comprend donc en quoi **le *machine learning* constitue à strictement parler une rupture par rapport à l'algorithmique classique.** Avec les algorithmes apprenants, c'est bien une nouvelle classe d'algorithmes qui émerge : on passe progressivement « d'un monde de programmation à un monde d'apprentissage » (Jean-Philippe Desbiolles, événement de lancement du débat public, CNIL, 23 janvier 2017). Les algorithmes classiques sont déterministes, leurs critères de fonctionnement sont explicitement définis par ceux qui les mettent en œuvre. Les algorithmes apprenants, au contraire, sont dits probabilistes. S'ils constituent une technologie bien plus puissante que les algorithmes classiques, leurs résultats sont mouvants et dépendent à chaque instant de la base d'apprentissage qui leur a été fournie et qui évolue elle-même au fur et à mesure de leur utilisation. Pour reprendre l'exemple du tigre (voir encadré), il est possible qu'une intelligence artificielle ayant été entraînée sur une base dans laquelle figure une seule espèce de tigres ne soit pas à même de reconnaître un tigre appartenant à une autre espèce. Mais on peut supposer qu'elle puisse aussi élargir sa capacité à reconnaître d'autres espèces de tigres à force d'être confrontée à de plus en plus d'individus partageant des traits communs aux deux races.

**Au-delà de ces différences techniques, une approche globale des algorithmes et de l'IA demeure cependant pertinente. Algorithmes déterministes et algorithmes apprenants soulèvent en effet des problèmes communs. Dans un cas comme dans l'autre, la finalité des applications de ces classes d'algorithmes consiste à automatiser des tâches autrement accomplies par des humains**, voire à déléguer à ces systèmes automatisés des prises de décisions plus ou moins complexes. Dès lors que l'on se détache d'une appréhension strictement technique de ces objets pour en aborder les conséquences et les implications sociales, éthiques, voire politiques, les problèmes posés se recoupent largement et méritent de faire l'objet d'une investigation conjointe.

Précisons enfin qu'algorithmes et intelligence artificielle recoupent à bien des égards ce que l'on appelle, de façon généralement imprécise, « Big data ». Le Big data désigne non seulement d'immenses quantités de données diverses mais également les techniques qui permettent de les traiter, de les faire parler, d'y repérer des corrélations inattendues, voire de leur conférer une capacité prédictive. De même, l'intelligence artificielle est indissociable des immenses quantités de données nécessaires pour l'entraîner et qu'elle permet en retour de traiter.

> L'algorithme sans données est aveugle. Les données sans algorithmes sont muettes

# Cadrer la réflexion en fonction des applications et des impacts les plus cruciaux des algorithmes aujourd'hui

En un sens l'algorithmique recouvre l'informatique et croise plus généralement tout ce qu'on a coutume d'englober sous le terme de « numérique ».

Face à un sujet potentiellement aussi vaste, il est donc aussi nécessaire que légitime de limiter le périmètre de la réflexion aux algorithmes qui posent aujourd'hui les questions éthiques et de société les plus pressantes. **La réflexion éthique sur les systèmes d'IA et sur les algorithmes n'a en effet de sens que si elle prend aussi en compte l'inscription de ceux-ci dans des contextes sociaux, humains, professionnels.**

Les pages qui suivent envisageront ainsi les usages de l'intelligence artificielle en limitant cette dernière aux usages s'appuyant sur le *machine learning,* qui sont les plus discutés aujourd'hui même si, en toute rigueur, ils ne constituent pas l'entièreté de ce domaine[6].

Par ailleurs, il a été décidé d'exclure du champ de la réflexion les problèmes soulevés par l'IA forte (ou générale). L'IA forte désigne des systèmes susceptibles de devenir complètement autonomes qui pourraient même se retourner contre l'homme. Cette vision se nourrit souvent d'un imaginaire apocalyptique alimenté par le cinéma hollywoodien dans le sillage de mythes parfois bien plus anciens (Frankenstein, etc.). Elle est souvent reliée à une interrogation concernant le niveau de conscience de soi d'une telle machine (en lien avec le thème de la singularité technologique). Elle est par ailleurs propagée par des prises de positions de personnalités du numérique disposant d'une forte visibilité médiatique, comme Elon Musk ou Stephen Hawking. Enfin, la diffusion du thème de la « singularité » par les milieux transhumanistes rayonnant depuis la Silicon Valley renforce les discours annonçant le dépassement prochain de l'homme par les machines. Force est pourtant de constater qu'elle est accueillie avec scepticisme par les plus éminents chercheurs et experts en informatique, comme en France Jean-Gabriel Ganascia. **L'hypothèse de l'avènement d'une IA forte est même dénoncée par certains (dont ce dernier) comme un moyen d'éluder de plus sérieux problèmes – éthiques voire tout simplement juridiques – que posent déjà et à brève échéance les progrès effectifs de l'IA faible et son déploiement croissant.**

Il aurait été possible, en toute rigueur et en prenant les termes au pied de la lettre, d'inclure dans le périmètre de notre réflexion sur les algorithmes les questions liées au chiffrement dans la mesure où cette technologie repose sur l'utilisation d'algorithmes. Le même procédé aurait pu conduire à considérer la « blockchain » comme partie intégrante du sujet. Là encore, il a semblé préférable d'adopter une attitude pragmatique, guidée par la perception publique de ce que sont aujourd'hui les algorithmes et leurs applications soulevant le plus de problèmes et d'interrogations. En d'autres termes, nous avons choisi de limiter le champ de la réflexion à ceux des algorithmes qui, dans l'immense diversité qui est la leur à l'ère numérique, soulèvent aujourd'hui des problèmes susceptibles d'interpeller directement le grand public et les décideurs, tant publics que privés.

À cet égard, les **algorithmes de recommandation,** s'ils ne constituent techniquement qu'une fraction des différents types d'algorithmes, constituent une partie importante de la question. Les algorithmes de recommandation sont employés pour établir des modèles prédictifs à partir d'une quantité importante de données et les appliquer en temps réel à des cas concrets. Ils élaborent des prévisions sur des comportements ou des préférences permettant de devancer les besoins des consommateurs, d'orienter une personne vers le choix jugé le plus approprié pour elle... Ces algorithmes peuvent par exemple être utilisés pour proposer des restaurants sur un moteur de recherche.

Si l'on prolonge cette approche, on peut lister ainsi les **principales fonctions et applications des algorithmes** susceptibles de faire débat et sur lesquelles la présente réflexion est centrée :

6 Les deux grandes approches de l'IA sont, d'une part, l'approche symboliste et cognitiviste et, d'autre part, l'approche neuro-inspirée et connexionniste (apprentissage automatique, réseaux de neurones, etc.). Les systèmes experts ont connu un important développement dans les années 1980. Les principales avancées récentes reposent sur l'apprentissage automatique.

- Produire des connaissances ;
- Apparier une demande et une offre (« matching »), répartir des ressources (passagers et chauffeurs de taxis, parents et places en crèche, étudiants et places à l'université, etc.) ;
- Recommander un produit, une offre de façon personnalisée ;
- Aider la prise de décision ;
- Prédire, anticiper (par exemple, des phénomènes naturels, des infractions, la survenue d'une maladie).

Ces grandes fonctions découlent de la capacité des algorithmes à filtrer l'information, à modéliser des phénomènes en identifiant des motifs parmi de grandes masses de données et à profiler les individus[7].

D'une manière générale, **la visibilité accrue des algorithmes et des questions qu'ils posent aujourd'hui est indissociable des masses de données inédites à disposition dans tous les secteurs qu'il faut trier pour pouvoir en tirer tout le potentiel.** La numérisation de notre société sous toutes ses formes – dématérialisation des transactions et services, révolution des capteurs, de l'Internet des objets, diffusion du smartphone, généralisation des politiques d'*open data,* etc. – est à l'origine de cette profusion. Celle-ci constitue aujourd'hui une ressource mais aussi un défi : **si nous avons besoin de recommandations, c'est que l'offre informationnelle est devenue pléthorique ; s'il est possible de profiler, c'est que la quantité de données collectées sur les individus permet de dépasser la segmentation par catégories prédéterminées.** L'enjeu soulevé par la qualité et la pertinence des données disponibles ou choisies pour alimenter les algorithmes constitue un autre point essentiel que rencontre toute réflexion à l'égard de ceux-ci.

Il faut aussi introduire l'idée d'**autonomisation,** pour bien prendre la mesure des enjeux soulevés par les algorithmes aujourd'hui. Si les algorithmes posent question, c'est aussi parce qu'ils permettent de déléguer des tâches auparavant accomplies par l'homme à des systèmes automatiques de plus en plus « autonomes ». Cependant, la délégation de tâches voire de décisions à des algorithmes traditionnels n'implique nullement que la production des algorithmes elle-même échappe à l'homme. L'intervention humaine est bien présente dans le recours aux algorithmes, par l'intermédiaire du paramétrage de l'algorithme, du choix et de la pondération des critères et des catégories de données à prendre en compte pour arriver au résultat recherché. Par exemple, si l'humain n'intervient pas directement dans la recommandation d'un restaurant par le biais d'une plateforme algorithmique, en revanche le rôle des développeurs est fondamental. En effet, ces derniers déterminent notamment l'importance que pourra jouer la localisation des restaurants, leur notation par d'autres usagers ou encore sa concordance supposée (là encore en fonction de critères à définir) avec le profil du requêteur.

Avec le développement du machine learning, on se situe un pas plus loin dans cette dynamique d'autonomisation, la machine écrivant « elle-même » les instructions qu'elle exécute, déterminant les paramètres qui doivent la guider dans le but d'accomplir une finalité qui reste cependant définie par l'homme.

> La visibilité accrue des algorithmes aujourd'hui est indissociable des masses de données inédites à disposition dans tous les secteurs, qu'il faut trier pour pouvoir en tirer tout le potentiel

7 Le profilage est défini par le Règlement européen sur la protection des données à caractère personnel comme « toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique ».

# Des usages et des promesses dans tous les secteurs

Les usages des algorithmes et de l'intelligence artificielle se développent dans tous les secteurs. Un discours porté très énergiquement par les acteurs économiques met en avant les avantages et les promesses de ces outils. On en mentionnera ici quelques exemples tout en renvoyant pour plus de détails aux **fiches sectorielles présentes en annexe et traçant les contours des grandes applications des algorithmes** que les débats ont permis d'évoquer[8].

Les usages aujourd'hui les plus banalisés ont trait, notamment, aux moteurs de recherche sur internet, aux applications de navigation routière, à la recommandation sur les plateformes de contenu culturel (type Netflix ou Amazon) ou sur les réseaux sociaux, au marketing pour le ciblage publicitaire et, de plus en plus, pour la prospection électorale.

Dans le domaine de la santé publique, l'utilisation des algorithmes est mise en avant pour la veille sanitaire (détection d'épidémies, de risques psycho-sociaux). On évoque de plus en plus les promesses d'une médecine de précision bâtissant des solutions thérapeutiques personnalisées en croisant les données du patient à celles de gigantesques cohortes.

Les fonctions régaliennes de l'État sont également concernées par l'émergence d'acteurs prétendant, par exemple, fournir des outils d'aide aux professions juridiques qui permettraient, à partir du traitement de données de jurisprudence, d'anticiper l'issue d'un procès ou d'affiner une stratégie judiciaire. Les services de police, en France et à l'étranger, commencent quant à eux à recourir à des outils algorithmiques destinés, par l'analyse de données, à orienter leurs efforts vers tel ou tel secteur.

Le débat largement médiatisé autour d' « APB » a mis en lumière aux yeux du grand public le recours à l'algorithme pour la répartition de centaines de milliers d'étudiants dans les universités. Au-delà de la gestion des flux, l'algorithme interroge les pratiques pédagogiques par des stratégies de personnalisation de l'enseignement toujours plus fines ou par la détection possible de décrochages scolaires.

**ENQUÊTE**

## Une connaissance inégale des usages des algorithmes*

L'intervention d'algorithmes est bien repérée par le public lorsqu'il s'agit d'un usage tel que le ciblage publicitaire (90 % des sondés en ont conscience).

Elle est en revanche souvent moins clairement perçue en ce qui concerne l'évaluation de la « compatibilité amoureuse » sur des applications de rencontre (46 % des répondants) ou l'élaboration d'un diagnostic médical (33 %).

* Enquête réalisée dans le cadre du débat public par l'association « Familles rurales », association familiale orientée vers les milieux ruraux, auprès de 1076 de ses adhérents.

Sur le marché de l'emploi, enfin, de nombreux acteurs travaillent actuellement au développement de solutions d'aide au recrutement (par appariement de l'offre et de la demande d'emploi, notamment) et de gestion des ressources humaines.

Sans prétendre épuiser un objet aux applications innombrables, le tableau qui figure à la page suivante donne cependant une idée de la façon dont les grandes fonctions identifiées des algorithmes et de l'intelligence artificielle se retrouvent dans différents secteurs.

8 Le développement industriel de l'intelligence artificielle est porté principalement par deux types d'acteurs. D'une part, des spécialistes de la fourniture de technologies et de services aux grandes entreprises, comme IBM avec Watson. D'autre part, les grands industriels de la donnée numérique (dont les GAFA), qui investissent fortement et chargent leurs services en IA (comme Google avec Translate, la reconnaissance d'images ou le traitement automatique de la parole).

## Les grandes fonctions des algorithmes et de l'IA dans différents secteurs

| | Education | Justice | Santé | Sécurité | Travail, RH | Culture | Autres |
|---|---|---|---|---|---|---|---|
| **Générer de la connaissance** | Mieux cerner les aptitudes d'apprentissage des élèves | Mettre en évidence les manières différenciées de rendre la justice selon les régions | Tirer profit de la quantité immense de publications scientifiques | Repérer des liens insoupçonnés pour la résolution d'enquêtes par les services de gendarmerie | Comprendre les phénomènes sociaux en entreprise | Créer des œuvres culturelles (peinture, musique) | Affiner le profil de risque d'un client d'un assureur |
| **Faire du matching** | Répartir les candidats au sein des formations d'enseignement supérieur (APB) | | Répartir des patients pour participation à un essai clinique | | Faire correspondre une liste de candidatures avec une offre d'emploi | | Mettre en relation des profils « compatibles » sur des applications de rencontres, etc. |
| **Prédire** | Prédire des décrochages scolaires | Prédire la chance de succès d'un procès et le montant potentiel de dommages-intérêts | Prédire des épidémies<br><br>Repérer des prédispositions à certaines pathologies afin d'en éviter le développement | Détecter les profils à risque dans la lutte anti-terroriste<br><br>Prédire l'occurrence future de crimes et délits | Détecter les collaborateurs qui risquent de démissionner dans les prochains mois | Créer des œuvres ayant un maximum de chance de plaire aux spectateurs (Netflix) | |
| **Recommander** | Recommander des voies d'orientation personnalisées aux élèves | Recommander des solutions de médiation en fonction du profil des personnes et des cas similaires passés | | | Proposer des orientations de carrière adaptées aux profils des personnes | Recommander des livres (Amazon), des séries télévisées (Netflix), etc. | Individualiser des messages politiques sur les réseaux sociaux |
| **Aider la décision** | | Suggérer au juge la solution jurisprudentielle la plus adéquate pour un cas donné | Suggérer au médecin des solutions thérapeutiques adaptées | Suggérer aux forces de police les zones prioritaires dans lesquelles patrouiller | | | Aider le conducteur à trouver le chemin le plus court d'un point à un autre (GPS) |

# Les enjeux éthiques

# Les enjeux éthiques

## L'éthique, éclaireuse du droit

La notion d'éthique fait souvent l'objet d'usages différents, laissant parfois place à une forme d'ambigüité. Les définitions proposées par les dictionnaires renvoient l'éthique à la morale, autrement dit à des normes qui n'ont pas nécessairement vocation à entrer dans le droit et qui portent sur la conduite des individus. Chez les philosophes antiques, l'éthique n'est ainsi rien d'autre que la réponse à la question suivante : « qu'est-ce qu'une vie bonne ? », c'est-à-dire des principes d'action qui concernent d'abord l'individu.

Plus récemment, la notion d'éthique s'est notamment développée comme renvoyant à une forme d'à côté du droit, évoqué entre autres par des acteurs privés comme les entreprises. L'éthique est alors un ensemble de normes édictées par l'entreprise et qu'elle s'impose à elle-même. Ces normes peuvent aller au-delà du droit. Souvent, elles peuvent n'avoir pour principale fonction que de redire – consciemment ou pas – des normes juridiques. Certaines évocations de l'utilisation « éthique » des données du client ne sont parfois rien d'autre qu'une façon de dire que l'entreprise se plie à la loi.

Un troisième usage de la notion d'éthique – sans doute le plus pertinent dans le contexte du présent rapport – s'est développé dans le langage des institutions publiques depuis la création en 1983 du Comité Consultatif National d'Ethique pour les sciences de la vie et de la santé (CCNE). Dans ce cadre, **l'éthique apparaît comme une éclaireuse du droit, la norme éthique une préfiguration de la norme juridique.** Que le législateur demande à une institution de produire une réflexion éthique place bien à l'horizon – plus ou moins proche – d'une telle réflexion l'inscription législative de celle-ci. La création par la loi du CCNE partageait un point commun important avec celle de la Loi pour une République numérique et sa création d'une mission de réflexion éthique confiée à la CNIL : un contexte marqué par de rapides avancées technologiques et par de fortes incertitudes sur l'attitude que la collectivité avait à adopter face à celles-ci. D'une part, les progrès de la biotechnologie (le premier bébé-éprouvette français naît en 1982), de l'autre

ENQUÊTE

### Une perception publique des algorithmes et de l'IA empreinte de méfiance*

Les trois craintes les plus partagées sont la **perte de contrôle humain** (63 % des adhérents), la **normativité** et l'enfermement à travers l'uniformisation des recrutements (56 %) et la **collecte disproportionnée de données personnelles** (50 %).

Dans le champ de l'emploi, quelques opportunités sont mises en exergue comme la possibilité d'examiner toutes les candidatures sur la base de critères identiques (52 %). Toutefois, **72 % des répondants envisagent comme une menace la possibilité d'être recruté par des algorithmes,** sur la base d'une analyse de leur profil et de sa compatibilité à un poste défini. 71 % d'entre eux affirment ainsi que la définition d'une charte éthique autour de l'usage des algorithmes constitue une réelle priorité.

**72 %**
**des répondants envisagent comme une menace la possibilité d'être recruté par des algorithmes**

* Enquête réalisée dans le cadre du débat public par la CFE-CGC, syndicat de l'encadrement, auprès de 1263 de ses adhérents (essentiellement issus des fédérations « Métallurgie » et « Finance et Banque »).

ce qui est ressenti comme une « révolution numérique ». L'inscription dans la loi d'une réflexion éthique répond donc au besoin d'un espace nécessaire pour une réflexion collective sur un pacte social dont certains aspects essentiels (libertés fondamentales, égalité entre les citoyens, dignité humaine) peuvent être remis en question dès lors que l'évolution technologique déplace la limite entre le possible et l'impossible et nécessite de redéfinir la limite entre le souhaitable et le non souhaitable.

La CNIL a choisi pour cette première réflexion de s'appuyer sur les acteurs désireux de s'exprimer sur les sujets liés aux algorithmes et à l'intelligence artificielle. Les enjeux éthiques retenus sont donc ceux qui ont été évoqués par ces mêmes acteurs. De manière logique, ces enjeux sont pour la plupart déjà bel et bien présents dans nos sociétés, même s'ils sont probablement appelés à gagner en intensité dans les années à venir. En revanche, des enjeux plus prospectifs, liés à des progrès pour l'heure hypothétiques des technologies numériques (transhumanisme, hybridation homme-machine, etc.), ont peu mobilisé la réflexion des partenaires impliqués et sont de ce fait peu développés dans le rapport.

> L'évolution technologique déplace la limite entre le *possible* et l'*impossible* et nécessite de redéfinir la limite entre le *souhaitable* et le *non souhaitable*

## LE REGARD DU CITOYEN

Les participants à la concertation citoyenne organisée par la CNIL à Montpellier le 14 octobre 2017 se sont prononcés sur les questions éthiques posées par les algorithmes et l'intelligence artificielle (voir « Une démarche innovante au service de l'élaboration d'une réflexion éthique collective et pluraliste ») : les enjeux qu'ils soulèvent résonnent en grande partie avec ceux identifiés tout au long du débat public.

Les citoyens semblent prioritairement préoccupés par les nouvelles modalités de prise de décision et la dilution de la responsabilité créées par l'algorithme. La « **perte de compétence** » **éventuelle de médecins ou d'employeurs** qui se reposeraient intensément sur l'algorithme a été mise en exergue. Parmi les conséquences préjudiciables évoquées : une « gestion des incertitudes » jugée inefficace chez la machine comparativement à ce dont est capable l'homme ; une incapacité à « gérer les exceptions » ou encore la « perte du sentiment d'humanité » (évoquées notamment à propos de l'absence de recours sur « APB »).

Le recours à des systèmes informatiques, parfois autonomes, pour prendre des décisions fait craindre que la **responsabilité** en cas d'erreurs ne soit « pas claire », une préoccupation soulevée notamment à propos du secteur médical. Concernant le cas « APB », certains citoyens critiquent le manque de transparence qui explique que l'algorithme serve « de bouc émissaire faisant tampon entre ceux qui font des choix politiques et ceux qui se plaignent de ces choix ». La problématique de la personnalisation informationnelle sur les réseaux sociaux et de ses effets collectifs, évoquée au sujet des élections présidentielles aux Etats-Unis, accentue également leur crainte que « plus personne ne soit réellement responsable du contrôle d'Internet ».

Moins évoqué, le danger de l'enfermement algorithmique est cependant mentionné par plusieurs participants des ateliers « ressources humaines » et « plateformes numériques ». Les citoyens ont aussi évoqué le risque de « **formatage** » des recrutements, et la rationalisation consécutive d'un champ qui ne devrait pas autant l'être, ou encore celui d'être figé sur Internet « dans un profil qui freinerait nos évolutions personnelles ».

Enfin, la thématique **des biais, des discriminations et de l'exclusion** mérite une vigilance toute particulière aux yeux des participants, et cela que les biais en question soit volontaires (en matière de recrutement, on craint l'éventualité qu'un algorithme soit codé « selon les objectifs des employeurs aux dépens des salariés ») ou involontaires (l'outil algorithmique est facteur d'inquiétudes quant aux erreurs qu'il pourrait générer).

# L'autonomie humaine au défi de l'autonomie des machines

Au-delà de la multiplicité des applications pratiques et des utilisations qui peuvent en être faites, **algorithmes et intelligence artificielle ont pour objet commun d'accomplir automatiquement une tâche ou une opération impliquant une forme d'« intelligence » qui serait autrement effectuée directement par un agent humain.** Autrement dit, il s'agit pour l'homme de **déléguer des tâches à des systèmes automatiques**[9].

**Le cas d'APB en offre un bon exemple.** Ce logiciel détermine l'affectation des bacheliers dans l'enseignement supérieur. Il peut être considéré comme ne faisant rien d'autre que d'appliquer un ensemble d'instructions et de critères qui pourraient tout aussi bien l'être par des fonctionnaires. L'intérêt essentiel du recours à l'algorithme est dans ce cas le gain de productivité induit par la délégation d'une tâche très coûteuse en temps et en moyens à un système automatique. Un autre intérêt de l'algorithme est de garantir le déploiement uniforme et impartial des règles définies en amont pour la répartition des futurs étudiants. En effet, l'application de ces mêmes règles par une chaîne administrative complexe peut donner prise, bien plus facilement, à des formes d'arbitraires ou même tout simplement à des interprétations différentes selon les agents qui les appliquent. Spécialiste des politiques éducatives, Roger-François Gauthier n'hésite ainsi pas à affirmer qu'APB a au moins eu le mérite de mettre fin à un système « mafieux » où le passe-droit avait sa place[10].

Si APB est un algorithme déterministe classique, l'utilisation de la **reconnaissance de formes pour identifier en temps réel des embarcations sur les photographies satellitaires** de très vastes surfaces maritimes fournit quant à elle une illustration de la façon dont l'intelligence artificielle permet aussi d'accomplir des tâches qui pourraient autrement s'avérer trop coûteuses en ressources humaines. Un simple logiciel peut ainsi assurer la surveillance 24 heures sur 24 de zones immenses qui nécessiterait autrement l'activité de nombreuses personnes.

De façon plus prospective, il serait au moins techniquement envisageable de confier – comme cela se fait déjà aux Etats-Unis – à des algorithmes le soin d'évaluer la dangerosité d'un détenu et donc l'opportunité d'une remise de peine. L'étape supplémentaire de ce que certains appellent la « justice prédictive » serait de confier à des systèmes le soin d'établir des décisions sur la base de l'analyse des données du cas à juger croisées aux données de jurisprudence.

## La délégation de tâches aux algorithmes : des situations contrastées

Il semble d'emblée assez évident que **les implications éthiques et sociales potentielles du phénomène accru de délégation de tâches à des systèmes automatisés présentent des degrés assez variés de sensibilité** selon les tâches qu'il s'agit de déléguer et selon les modalités mêmes de cette délégation.

Il est ainsi possible de faire un pas supplémentaire pour distinguer les cas sur lesquels la réflexion doit se concentrer, au moyen d'**une typologie du phénomène de délégation d'opérations à des systèmes automatisés,** en fonction de deux critères : **l'impact sur l'homme de l'opération qu'il s'agit de déléguer et le type de système à qui il est question de déléguer celle-ci.**

Le premier critère concerne le type d'impact et/ou l'ampleur de l'opération déléguée au système automatisé. Il peut s'agir d'une tâche routinière, mécanique et relativement anodine (par exemple, le classement par ordre alphabétique d'une série de fichiers informatiques). À l'opposé, cette tâche peut perdre son caractère anodin et s'avérer d'une grande complexité. Elle peut, surtout, prendre les aspects d'une *décision* et revêtir une importance vitale pour une personne ou pour un groupe, comme lorsqu'il s'agit d'établir une aide au diagnostic médical. Entre ces deux extrêmes se déploie un large spectre de situations contrastées. On y retrouverait les deux exemples évoqués ci-dessus ou encore celui de la voiture autonome, ce dernier ainsi que le cas d'APB étant relativement plus proches du cas du diagnostic médical automatisé que de l'autre bout du spectre.

Le second critère concernerait quant à lui le type de système automatisé – algorithme classique ou algorithme de *machine learning* – à qui l'on délègue l'opération. Une autre façon de présenter ce critère est d'évoquer le **degré d'autonomie du système** en question, en particulier sa capacité ou non à élaborer ses propres critères de fonctionnement. De même, ce critère renvoie à la capacité ou non du système de produire une explication satisfaisante des résultats qu'il fournit.

Cette typologie souligne la grande diversité des situations impliquées par une réflexion sur les enjeux éthiques et sociaux des algorithmes et de l'intelligence artificielle. Elle met surtout en évidence l'étendue du spectre sur lequel peut

9 En toute rigueur, rappelons-le, ce n'est d'ailleurs généralement pas tant le recours à l'algorithme qui constitue le fait nouveau que son exécution sous la forme d'un programme informatique.
10 Événement de lancement du débat public, CNIL, 23 janvier 2017.

se situer le degré de gravité ou de sensibilité des enjeux liés à l'utilisation de tel ou tel algorithme.

## La délégation de décisions critiques aux algorithmes: une déresponsabilisation ?

Les décisions les plus cruciales (diagnostics médicaux, décisions judiciaires, décision d'ouvrir le feu dans un contexte de conflit armé etc.) qui pourraient être, voire commencent à être (à l'étranger notamment) déléguées à des systèmes automatisés sont – au moins dans certains cas – déjà clairement thématisées par la tradition juridique, en France. Seul un médecin est ainsi habilité à établir un diagnostic qui, autrement, relèverait de l'exercice illégal de la médecine. Il en va de même de la décision du juge, qui ne saurait en toute rigueur être déléguée à un système automatisé. Dans cette perspective, ce type de système est présenté dans ces domaines comme une « aide » à la prise de décision.

Cette clarté juridique ne résout cependant pas les problèmes que soulève l'éventualité d'une délégation de ce type de décisions. Comment s'assurer que la prédiction et la recommandation fournies par les algorithmes ne soient effectivement qu'une aide à la prise de décision et à l'action humaine sans aboutir à une déresponsabilisation de l'homme, à une perte d'autonomie ?

Dans le domaine médical où la qualité de la prise de décision peut être plus facilement évaluée (ou, du moins, quantifiée), on peut logiquement se demander quelle marge d'autonomie resterait au médecin face à la recommandation (en termes de diagnostic et de solution thérapeutique à privilégier) qui serait fournie par un système d'« aide » à la décision extrêmement performant. On annonce en effet que l'intelligence artificielle serait supérieure à l'homme pour le diagnostic de certains cancers ou pour l'analyse de radiographies. Dans le cas où ces annonces s'avéreraient exactes, il pourrait donc devenir hasardeux pour un médecin d'établir un diagnostic ou de faire un choix thérapeutique autre que celui recommandé par la machine, laquelle deviendrait dès lors le décideur effectif. Dans ce cas, se pose alors la question de la responsabilité. Celle-ci doit-elle être reportée sur la machine elle-même, qu'il s'agirait alors de doter d'une personnalité juridique ? Sur ses concepteurs ? Doit-elle être encore assumée par le médecin ? Mais alors, si cela peut certes sembler résoudre le problème juridique, cela n'aboutit-il quand même pas à une déresponsabilisation de fait, au développement d'un sentiment d'irresponsabilité ?

FOCUS

## Les défis éthiques d'une police prédictive

La quête d'une prédiction du crime dans le temps et dans l'espace serait capable de prédire le crime dans le temps et dans l'espace, afin d'orienter l'action des patrouilles, fait l'objet d'un développement actif de logiciels algorithmiques. Aux Etats-Unis, « **PredPol** » s'appuie sur des modèles empruntés à la sismologie pour évaluer l'intensité du risque à tel endroit et à tel moment. La start-up prétend ainsi intégrer la dimension « contagieuse » de la diffusion spatiotemporelle des délits.

Ce potentiel prédictif s'est pourtant révélé limité, d'une part, car la contagion a un impact négligeable pour la détection de crimes comparativement aux répliques d'un séisme et, d'autre part, car la structure de la criminalité varie d'une année à l'autre. Pourtant, cela ne dissipe pas l'attrait de tels dispositifs consistant à permettre de ***« gérer, selon des critères gestionnaires, l'offre publique de vigilance quotidienne ».*** Très concrètement, *« le carré prédictif reste rouge sur la carte tant que la police n'y a pas patrouillé, il tourne ensuite au bleu lors des premiers passages, puis il apparaît en vert lorsque le policier a passé le temps suffisant et optimal calculé selon les ressources disponibles »*[11].

Une crainte majeure émerge : quid du risque que les préconisations de la machine soient appréhendées comme une vérité absolue, non soumise à la discussion quant à ses conséquences pratiques ? Dans la mesure où l'algorithme se repose sur les données issues des plaintes des victimes, une conséquence pratique constatée est celle d'une présence policière renforcée dans les zones où les publics portent plainte avec plus de fluidité, et ainsi un phénomène d'exclusion de l'offre de sécurité publique pour certaines populations (celles qui signalent moins). On peut imaginer, au contraire, que l'utilisation de ce type d'algorithme focalise l'attention policière sur certains types d'infractions au détriment d'autres.

Dans tous les cas, une appréhension critique de ce type d'outil est une nécessité majeure. Quid également de la capacité à juger de l'efficacité de ces modèles ? Qu'un délit soit détecté par une patrouille orientée par le système, ou que ce ne soit pas le cas, le résultat pourrait facilement (mais faussement) être interprété comme un signe de l'efficacité de l'outil.

11 Bilel Benbouzid, « A qui profite le crime ? Le marché de la prédiction du crime aux Etats-Unis », www.laviedesidees.fr

Le cas de la médecine est particulièrement critique non seulement en raison de l'impact des décisions et recommandations sur les personnes mais aussi en raison du fait que la discussion implique ici des systèmes fondés sur la technologie du *machine learning*. Ceci implique que les logiques sous-jacentes des systèmes d'intelligence artificielle sont potentiellement incompréhensibles pour celui à qui ils sont proposés, autant d'ailleurs que pour les concepteurs du système. Le débat public organisé par la CNIL a d'ailleurs été l'occasion de constater une controverse sur ce point, à propos notamment du logiciel Watson d'IBM. Le discours d'IBM souligne que Watson fonctionne sur le mode de l' « apprentissage supervisé ». Autrement dit, le système est accompagné pas à pas dans son apprentissage, ce qui permettrait d'en contrôler la logique, par opposition à un apprentissage non supervisé qui reviendrait effectivement à laisser une pleine et entière autonomie à la machine pour déterminer ses critères de fonctionnement. IBM indique également contrôler le fonctionnement des systèmes avant de décider de conserver l'apprentissage réalisé. Au contraire, les chercheurs experts de ce domaine qui ont eu l'occasion de s'exprimer lors des différents débats organisés (et notamment la CERNA) ont régulièrement rappelé qu'en l'état actuel de la recherche les résultats fournis par les algorithmes de machine learning les plus récents n'étaient pas explicables. Cette explicabilité constitue d'ailleurs l'objet de recherches en cours. Ils insistent également sur le fait qu'il est très difficile de contrôler effectivement un système de *machine learning*.

**On peut ainsi se demander si les algorithmes et l'intelligence artificielle ne conduisent pas à une forme de dilution de figures d'autorité traditionnelles, de décideurs, de responsables, voire de l'autorité même de la règle de droit.** Cette évolution est parfois explicitement souhaitée. Certains, comme Tim O'Reilly, imaginent d'ores et déjà l'avènement d'une « réglementation algorithmique[12]» qui verrait la « gouvernance » de la cité confiée aux algorithmes : grâce aux capteurs connectés, lieux, infrastructures et citoyens communiqueraient en permanence des données traitées en vue de rationaliser et d'optimiser la vie collective selon des lois considérées comme « naturelles », émanant des choses mêmes, une « normativité immanente », comme l'expliquent Thomas Berns et Antoinette Rouvroy[13]. Sans doute faut-il remarquer ici que la tentation – révélée par ces discours – de se passer d'une normativité humaine et de préférer une normativité algorithmique est favorisée par les discours marchands. Ces derniers vantent l' « objectivité » supposée des systèmes automatiques (par opposition à un jugement humain toujours faillible). Ils influent donc sur la tendance des utilisateurs à prendre le résultat produit par une machine pour une vérité incontestable, alors même qu'il est de part en part déterminé par des choix (de critères, de types de données fournies au système) humains[14].

L'impact des algorithmes sur la conception et l'application de la norme pourrait aussi prendre une autre forme. Le Conseil National des Barreaux, dans le rapport qu'il a remis à la CNIL, souligne ainsi qu'**« il faut éviter que l'obsession de l'efficacité et de la prévisibilité qui motive le recours à l'algorithme nous conduise à concevoir les catégories et les règles juridiques non plus en considération de notre idéal de justice mais de manière à ce qu'elles soient plus facilement « codables » ».**

Il n'est pas exclu que cette évolution progressive vers des formes de « réglementation algorithmique » puisse présenter une sorte d'attrait pour les décideurs eux-mêmes. Déléguer des décisions à une machine – supposée neutre, impartiale, infaillible – peut être une façon d'éluder sa propre responsabilité, de s'exempter de la nécessité de rendre compte de ses choix. Le **développement d'armes létales autonomes (robots tueurs)** qui pourraient prendre elles-mêmes la décision de tuer sur le champ de bataille ou à des fins de maintien de l'ordre soulève la question avec une particulière acuité. L'acte de tuer, même considéré comme légitime, dans une situation de conflit international et face à un ennemi armé, ne doit-il pas rester sous le contrôle et la responsabilité directe de l'homme ? Sa difficulté et son caractère éventuellement traumatique pour celui-là même qui l'accomplit ne doivent-ils pas être considérés comme une garantie nécessaire pour éviter toute dérive ?

**Ces considérations ne concernent pas que les situations où des tâches ou des décisions sont déléguées à un algorithme apprenant. L'algorithme classique, déterministe, est également concerné.** Les débats autour de l'algorithme d'APB en ont offert un bon exemple, sinon une manière de comprendre comment **peut se mettre en place un tel processus de dépolitisation et de neutralisation de choix de société méritant pourtant de faire l'objet d'une discussion publique.** La polémique s'est en effet concentrée sur l'algorithme lui-même, notamment à la suite de la révélation de la mise en œuvre du tirage au sort qu'il induisait pour certains candidats à des filières en tension. Or, l'algorithme n'est jamais que le reflet de choix politiques, de choix de société. En l'occurrence, le recours au tirage au sort pour l'attribution de places dans des filières en tension est le résultat d'un choix politique dont deux alternatives possibles seraient – schématiquement – la sélection à l'entrée à l'université ou l'investissement pour faire correspondre le nombre de places disponibles dans les filières en question avec la demande. En d'autres termes, « code is law », pour reprendre la fameuse formule de Lawrence Lessig.

12 Tim O'Reilly, « Open data and algorithmic regulation », in Brett Goldstein (dir.), Beyond Transparency: Open Data and the Future of Civic Innovation, San Francisco, Code for America, 2013, pp. 289-301.
13 Rouvroy Antoinette, Berns Thomas, « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ? », Réseaux, 2013/1 (n° 177), p. 163-196.
14 La prétendue objectivité machinique n'est à ce titre qu'une subjectivité diluée et non assumée.

On ne saurait en effet considérer qu'un algorithme (entendu au sens large comme le système socio-technique dont il fait partie) puisse être « neutre », dans la mesure où il incorpore inévitablement des partis pris – que ceux-ci soient sociaux, politiques, éthiques ou moraux – et répond le plus souvent à des finalités qui incluent une dimension commerciale pour son auteur. L'exemple fréquemment évoqué du choix que pourrait être amené à faire l'algorithme d'une voiture sans chauffeur de sacrifier ou bien son occupant ou bien un piéton sur la route illustre la façon dont le recours à la technique, plus que de soulever certains problèmes moraux, a surtout pour effet de les déplacer : à un dilemme réglé en temps réel par une personne impliquée dans sa chair fait place un choix effectué par d'autres, ailleurs, bien en amont[15].

Au-delà de la finalité délibérément visée à travers la mise en place d'APB (efficacité administrative renforcée et harmonisation plus équitable de l'attribution de places dans l'enseignement supérieur), force est de constater que celle-ci a pour effet induit l'escamotage de choix de société impliqués par le paramétrage du système mais masqués par l'impartialité supposée de l'algorithme. Les responsables de la mise en œuvre de l'algorithme auquel est déléguée une prise de décision devraient donc chercher des moyens de contrer ce type d'effets (par exemple, par un effort d'information du public concerné). Ils devraient en tout cas s'interdire de l'exploiter en se cachant derrière la machine ou même de s'en accommoder dans la mesure où il a tendance à neutraliser des conflits ou des débats légitimes.

> Les algorithmes et l'intelligence artificielle conduisent à une forme de dilution de figures d'autorité traditionnelles, de décideurs, de responsables, voire de l'autorité même de la règle de droit

Il est d'ailleurs probable que céder à cette facilité ait pour contrepartie un sentiment d'inhumanité chez les personnes concernées. Ce sentiment est susceptible de se transformer en rejet, en particulier si n'est prévue aucune possibilité de contacter l'organisme responsable et d'échanger pour « trouver des solutions ou tout simplement pour être écouté », ainsi que l'a souligné le médiateur de l'Éducation nationale[16].

Dans le cas d'un algorithme déterministe tel qu'évoqué ici, la dilution de la responsabilité n'est pourtant qu'apparente. Les choix et les décisions cruciales se trouvent tout simplement déplacés au stade du paramétrage de l'algorithme.

Est-ce à dire que ceux qui maîtrisent le code informatique deviennent les véritables décideurs et que se profile le risque que le pouvoir se trouve concentré dans les mains d'une « petite caste de scribes » (Antoine Garapon, événement de lancement du débat, le 23 janvier 2017) ? Ce n'est certes pas ce qu'a donné à voir le cas d'APB. Suite à l'ouverture du code source des algorithmes de l'administration qu'a imposée la loi pour une République numérique, celui d'APB a été examiné par la mission Etalab. Il s'est avéré que ses développeurs avaient pris soin d'y documenter l'origine de chaque modification du paramétrage de l'algorithme, en l'occurrence les directives qu'ils avaient reçues de la part de l'administration. En somme, la traçabilité de la responsabilité a été organisée par les développeurs mêmes d'APB. Cet exemple ne doit cependant pas masquer le fait que la logique algorithmique a tendance à déporter la prise de décision vers les étapes techniques de conception d'un système (paramétrage, développement, codage), lequel ne fait ensuite que déployer automatiquement et sans faille les choix opérés initialement. La préoccupation d'Antoine Garapon évoquée précédemment ne saurait donc pas être écartée et appelle des réponses. **Il est essentiel que ces étapes de conception ne s'autonomisent pas exagérément au point de devenir le lieu de la prise de décision.**

**La question du lieu de la responsabilité et de la décision se pose en partie différemment dès lors qu'il s'agit de systèmes de *machine learning*.** Sans doute faut-il ici davantage penser en termes de chaîne de responsabilité, depuis le concepteur du système jusqu'à son utilisateur, en passant par celui qui va entraîner ce système apprenant. En fonction des données qui lui auront été fournies, ce dernier se comportera différemment, en effet. On peut penser ici au cas du robot conversationnel Tay mis en place par Microsoft et suspendu au bout de 24 heures quand, alimenté par des données d'utilisateurs des réseaux sociaux, il avait commencé à tenir des propos racistes et sexistes. Reste qu'organiser précisément la répartition de la responsabilité entre

**15** Voir à ce sujet l'excellent site du MIT offrant une illustration pratique de ces dilemmes : http://moralmachine.mit.edu/
**16** Le Monde, 29 juin 2016 : « Le médiateur de l'Education Nationale dénonce la solitude des familles face à APB ».

ces différents maillons de la chaîne est un problème ardu. Au-delà, **faut-il conditionner l'utilisation de l'intelligence artificielle à la capacité d'attribuer de façon absolument claire cette responsabilité ?** On sait d'ores et déjà que des intelligences artificielles peuvent être plus « performantes » que l'homme pour réaliser certaines tâches, sans que l'on ait une claire compréhension du fonctionnement de ces systèmes et donc, aussi, des erreurs éventuelles qu'ils pourraient commettre. Rand Hindi explique ainsi que « les IA font moins d'erreurs que les humains mais qu'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 ». Faut-il alors imaginer d'attribuer une personnalité juridique à ces systèmes ? Ou faire endosser la responsabilité à l'utilisateur lui-même (en l'occurrence, dans le domaine médical, au patient) ?

Sans doute ne faut-il toutefois pas exagérer la spécificité du cas du *machine learning*. Imaginons une intelligence artificielle chargée de répartir les malades dans les services d'un hôpital et de fixer la fin de leur hospitalisation de la manière la plus « efficace » possible. Certainement, le système aurait une part d'opacité liée à son caractère apprenant. Mais, dans le même temps, les objectifs qui lui seraient assignés, ainsi que leur pondération (garantir le maximum de guérisons à long terme, minimiser le taux de réhospitalisations à brève échéance, rechercher la brièveté des séjours, etc.), seraient bien des choix explicitement faits par l'homme.

## Une question d'échelle : la délégation massive de décisions non critiques

La réflexion éthique sur les algorithmes et l'intelligence artificielle doit-elle se cantonner à considérer les décisions cruciales, les secteurs où l'impact sur l'homme est incontestable, comme la médecine, la justice, l'orientation scolaire, voire l'automobile, avec ses implications en termes de sécurité ? **Ne faut-il pas prendre en compte également les algorithmes à qui nous sommes amenés à déléguer progressivement de plus en plus de tâches et de décisions apparemment anodines mais qui, mises bout à bout, constituent l'étoffe de nos existences quotidiennes ?**

Simplement par leur capacité à fonctionner de façon répétée, sur de longues durées et surtout à de très vastes échelles, les algorithmes peuvent avoir un impact considérable sur les personnes ou sur les sociétés. Par exemple, les critères de fonctionnement d'une banale application de guidage automobile, dès lors qu'ils sont utilisés par un nombre conséquent d'automobilistes qui s'en remettent implicitement à eux pour décider des itinéraires qu'ils empruntent, peuvent avoir des impacts importants sur le trafic urbain, la répartition de la pollution et à terme, peut-être, sur la forme même de la ville et de la vie urbaine. Le Laboratoire d'innovation numérique (LINC) de la CNIL l'explique ainsi : « Hormis la question de la captation des données personnelles, se pose celle de la perte de contrôle de l'acteur public sur l'aménagement de l'espace public, sur la gestion des flux, et au-delà sur la notion même de service public et d'intérêt général. La somme des intérêts individuels des clients d'un Waze peut parfois entrer en contradiction avec les politiques publiques portées par une collectivité[17] ».

Cathy O'Neil, dans son ouvrage *Weapons of Math Destruction*[18], propose un exemple particulièrement évocateur. Elle imagine qu'elle pourrait modéliser les règles qu'elle suit implicitement pour composer les repas de ses enfants (diversité, présence de légumes verts mais dans des limites permettant de prévenir de trop fortes protestations, relâchement des règles les dimanches et jours de fête, etc.). Un programme mettant en œuvre un tel algorithme ne poserait pas de problème tant qu'il ne serait utilisé pour générer automatiquement des repas que pour un nombre limité de personnes. Or, la caractéristique spécifique des algorithmes exécutés par des programmes informatiques est leur échelle d'application. Un tel programme, utilisé tel quel par des millions de personnes, aurait nécessairement des impacts puissants et potentiellement déstabilisateurs sur de grands équilibres sociaux et économiques (renchérissement du prix de certaines denrées, effondrement de la production d'autres produits, uniformisation de la production, impact sur les professions de la filière agro-industrielle, etc.). **C'est ici un aspect bien spécifique des algorithmes informatiques déployés aujourd'hui à l'heure d'Internet qui constitue le fait nouveau et que l'auteur met en évidence : leur *échelle* de déploiement.** Sans doute cet aspect ne saurait-il être ignoré par ceux qui déploient des algorithmes susceptibles d'être utilisés à une large échelle.

## L'optimisation algorithmique comme écrasement du temps et de l'espace

L'une des caractéristiques du fonctionnement algorithmique est son immédiateté et sa simplicité, du moins son uniformité et son caractère inexorable. Les algorithmes d'IA ont la capacité d'accomplir une tâche dans un temps presque immédiat (réduit au temps du seul calcul de la machine). Ils ont la capacité d'accomplir cette même tâche à une très large échelle spatiale mais de façon identique en tous lieux. À ce titre, ils peuvent présenter un grand attrait pour des administrations ou des entreprises soucieuses d'efficacité mais aussi de rationalité et d'homogénéité de leur action.

17 CNIL (LINC), La Plateforme d'une ville. Les données personnelles au cœur de la fabrique de la smart city, Cahier IP n°5, octobre 2017, p. 20.
18 Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown, 2016.

Or, cette caractéristique des algorithmes implique aussi une dimension potentiellement problématique : **l'écrasement de la durée et de la dimension spatiale du processus délégué à la machine peut aussi constituer une perte, un appauvrissement de l'action.** Les cas des algorithmes utilisés par l'administration ainsi que celui de la justice prédictive permettent de mieux saisir cette ambivalence, entre optimisation et appauvrissement de processus vidés de leur dimension spatiale.

Ainsi, le déploiement d'un algorithme comme celui du logiciel APB peut certes être considéré comme garant pour l'administration d'une forme de simplicité et d'harmonisation de l'application des règles, là où le fonctionnement d'une chaîne administrative complexe et nombreuse peut donner prise à des différences d'interprétation et d'application. Pourtant, ce qui peut apparaître à première vue comme un manque d'efficacité ou comme le signe d'un fonctionnement parfois erratique ne peut-il pas être aussi considéré comme une source précieuse d'information pour les décideurs, via les retours d'expériences et les questionnements de ceux qui sont chargés d'appliquer les règles et peuvent en observer le déploiement et éventuellement les limites, au plus près du terrain ?

De même, le colloque sur la justice prédictive organisé le 19 mai 2017 par le Barreau de Lille, la Faculté de droit de l'Université catholique de Lille et la cour d'appel de Douai a vu certains participants souligner que « la connaissance des décisions rendues par les autres juridictions voisines ou par les autres magistrats contribuera à une certaine harmonie et évitera que l'issue d'un litige dépende de la question de savoir s'il est plaidé à Lille ou à Marseille ». L'idée repose ici sur la capacité des algorithmes à traiter les grandes masses de données de jurisprudence mises en open data et à mettre en évidence des disparités d'application de la loi dans différentes juridictions. Le dévoilement de ces disparités dont le juge n'a pas lui-même conscience aurait pour conséquence une harmonisation de l'application de la loi sur le territoire national. Pourtant, est-on absolument certain que, dans certaines limites, des formes de disparités régionales ne traduisent pas en fait un usage raisonné de la prudence du juge et l'adaptation intelligente et fine de celui-ci à des réalités sociales pouvant varier d'un lieu à l'autre ? Une forme de respiration de la loi, peut-être, à distinguer de son application automatique et rigide ?

On peut appliquer le même type de raisonnement à l'idée d'une justice prédictive qui, poussée à son extrême (une décision de justice rendue par une intelligence artificielle), éluderait l'apport de la délibération en commun et de ce qui peut s'y jouer à travers la confrontation d'individualités partageant un objectif commun. **La délibération de jurés et de magistrats n'est pas que le simple déploiement d'arguments préexistants à la manière dont un logiciel « exécute » un programme. La durée n'y est pas qu'un décor accessoire, une ressource dont il conviendrait de limiter la dépense : elle y est un acteur à part entière.** Elle implique la capacité des jurés à évoluer au cours de l'échange d'arguments, à changer de positions, ainsi que le montre mieux que toute démonstration le film de Sidney Lumet, *Douze hommes en colère*.

Il semble en tout cas souhaitable d'attirer l'attention des utilisateurs d'algorithmes et d'intelligence artificielle sur la nécessité de ne pas prendre en compte seulement les apports, mais aussi les inconvénients éventuels de ces technologies, leur caractère potentiellement ambivalent, et de réfléchir aux moyens de les contrer.

# Biais, discriminations et exclusion

La propension des algorithmes et de l'intelligence artificielle à générer des biais pouvant conduire à leur tour à créer ou à renforcer des discriminations s'est imposée comme un sujet d'inquiétude et de questionnement. Le constat mérite d'autant plus d'être souligné que ces systèmes techniques peuvent également parfois nourrir une croyance en leur objectivité. Une objectivité d'autant plus précieuse qu'elle ferait souvent défaut aux humains. Tout algorithme est pourtant, en un sens, biaisé, dans la mesure où il est toujours le reflet – à travers son paramétrage et ses critères de fonctionnement, ou à travers les données d'apprentissage qui lui ont été fournies – d'un système de valeurs et de choix de société. Le débat autour des biais et des discriminations qu'ils peuvent générer n'est donc qu'un miroir grossissant mettant en valeur cette caractéristique essentielle dans ce qu'elle a de plus problématique.

Plusieurs exemples ont récemment illustré de façon particulièrement nette et choquante ce type de biais. En 2015, un logiciel de reconnaissance faciale de Google a ainsi suscité une forte polémique. Un jeune couple d'Afro-Américains s'est rendu compte qu'une de ses photos avait été

étiquetée sous le tag « gorille ». L'explication de ce dysfonctionnement réside dans le type de données avec lesquelles l'algorithme a été entraîné pour reconnaître des personnes. En l'occurrence, il est vraisemblable qu'il l'ait été au moyen essentiellement, voire exclusivement, de photographies de personnes blanches (d'autres exemples existent d'ailleurs de biais racistes de logiciels de reconnaissance d'image au détriment de personnes de type « asiatique »). En conséquence, l'algorithme a considéré qu'une personne de couleur noire présentait plus de similitude avec l'objet « gorille » qu'elle avait été entrainée à reconnaitre qu'avec l'objet « humain ».

Notons d'ailleurs que des actes de malveillance volontaires de la part de personnes impliquées dans le processus d'entraînement de ce type d'algorithmes ne sont pas exclus. Ainsi en a-t-il été pour le robot conversationnel Tay développé par Microsoft et qui s'est mis à proférer sur Twitter des propos racistes et sexistes après quelques heures de fonctionnement et d'entraînement au contact des propos que lui adressaient des internautes.

Les biais des algorithmes peuvent aussi être des biais de genre. En 2015, trois chercheurs de l'Université Carnegie Mellon et de l'International Computer Science Institute ont mis en évidence la façon dont Adsense, la plateforme publicitaire de Google, générait un biais au détriment des femmes. À l'aide d'un logiciel baptisé Adfisher, ils ont créé 17 000 profils dont ils ont ensuite simulé la navigation sur le Web afin de mener une série d'expériences. Ils ont ainsi constaté que **les femmes se voyaient proposer des offres d'emploi moins bien rémunérées que celles adressées à des hommes, à niveau similaire de qualification et d'expérience.** Il est apparu qu'un nombre restreint de femmes recevaient des annonces publicitaires en ligne leur proposant un emploi au revenu supérieur à 200 000 dollars annuels. Loin d'être anecdotique, « la publicité en ligne ciblée de Google est tellement omniprésente que l'information proposée aux personnes est susceptible d'avoir un effet tangible sur les décisions qu'elles prennent », souligne Anupam Datta, co-auteur de l'étude.

Ici encore, les causes précises sont difficiles à établir. Il est bien sûr envisageable qu'un tel biais soit le fruit d'une volonté des annonceurs eux-mêmes : ceux-ci auraient alors délibérément choisi d'adresser des offres différentes aux hommes et aux femmes. Mais il est tout aussi possible que ce phénomène soit aussi le résultat d'une réaction de l'algorithme aux données qu'il a reçues. En l'occurrence, les hommes auraient pu avoir davantage tendance en moyenne à cliquer sur les publicités annonçant les emplois les mieux rémunérés tandis que les femmes auraient eu tendance à s'autocensurer, selon des mécanismes bien connus des sciences sociales. Dès lors, le biais sexiste de l'algorithme ne serait pas autre chose que la reproduction d'un biais préexistant dans la société.

FOCUS

## Des algorithmes contre la récidive ?

Les applications de justice prédictive font l'objet d'une attention publique toute particulière quant à leurs potentiels effets discriminatoires. Une polémique a éclaté autour de l'application COMPAS (Correctional Offender Management Profile for Alternative Sanction) visant à produire un **score de risque de récidive** pour les détenus ou accusés lors d'un procès. Bien que des outils d'analyse statistique de données aient déjà été déployés au sein des tribunaux américains depuis les années 1970, un tel calcul automatique sous la forme de score revêt un caractère nouveau pour la prise de décisions de libération conditionnelle.

En d'autres termes, le travailleur social utilisant COMPAS a recours à une interface lui permettant de répondre, en collaboration avec le prévenu, à des questions du type « Que pense le prévenu de la police ? », « Quelles sont les caractéristiques des amis du prévenu ? », « Certains d'entre eux ont-ils déjà été condamnés ? »[19]. Un score de risque est ainsi calculé et ajouté au dossier du prévenu.

Le site ProPublica a accusé Nortphointe, société commercialisant COMPAS, de produire des scores **biaisés et racistes**[20]**.** Ce constat repose sur la confrontation des scores de récidive de détenus libérés avec l'observation, ou non, d'une arrestation sur une période de deux ans. Le taux de faux positifs (c'est-à-dire un score élevé mais sans récidive effective observée) s'est révélé considérablement plus fort pour les anciens détenus d'origine afro-américaine que pour les individus blancs.

19 https://usbeketrica.com/article/un-algorithme-peut-il-predire-le-risque-de-recidive-des-detenus
20 https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Troisième exemple, en avril 2016, il a été révélé qu'Amazon avait exclu d'un de ses nouveaux services (la livraison gratuite en un jour) des quartiers peuplés majoritairement de populations défavorisées à Boston, Atlanta, Chicago, Dallas, New York et Washington. À l'origine, un algorithme d'Amazon avait mis en évidence, en analysant les données à sa disposition, que les quartiers en question n'offraient guère de possibilités de profit pour l'entreprise. Même si l'objectif d'Amazon n'était assurément pas d'exclure de ses services des zones parce que leur population était majoritairement noire, tel s'avérait pourtant bien être le résultat de l'utilisation de cet algorithme : dans six grandes villes, il apparaît clairement que « l'aire de fourniture du service exclut les codes postaux à population majoritairement noire, à des degrés variés ». En conséquence, les citoyens noirs ont environ deux fois moins de chances que les blancs de vivre dans des zones desservies par [le service d'Amazon en question][21] ». À Boston, alors que la ville entière avait accès au service, seuls trois codes postaux en étaient exclus, dans le quartier majoritairement noir de Roxbury.

Comment expliquer ce phénomène, alors qu'Amazon a souligné – à juste titre, sans aucun doute – n'avoir recouru à aucune donnée raciale pour alimenter l'algorithme ? Il a été opposé à Amazon que les quartiers concernés étaient précisément les mêmes que ceux qui avaient fait l'objet pendant des décennies de la pratique dite du « redlining », consistant pour les banques à refuser systématiquement d'accorder des prêts à des Afro-Américains, même solvables, en raison de la couleur de leur peau et de leur domiciliation dans des zones peuplées majoritairement par des minorités. Il est donc évident que l'algorithme d'Amazon a pour effet de reproduire des discriminations préexistantes, quand bien même aucun racisme intentionnel n'est ici à l'œuvre.

Le paramétrage des algorithmes, c'est-à-dire la définition explicite des critères selon lesquels ils fonctionnent et opèrent des tris, sélectionnent et recommandent, peut bien sûr être la source de biais et de discrimination. Mais, comme le montrent les trois exemples évoqués ci-dessus, ce sont bien les biais provoqués par les données fournies aux systèmes qui soulèvent le défi le plus redoutable aujourd'hui. Le caractère historique d'un jeu de données confère à celui-ci la capacité à reproduire des inégalités ou des discriminations préexistantes. Un algorithme qui chercherait à définir les profils à recruter sur la base des profils ayant correspondu aux trajectoires de carrière les plus réussies dans le passé d'une entreprise pourrait ainsi tout à fait exclure les femmes, soit que celles-ci aient fait l'objet d'une exclusion dans le passé, soit qu'elles aient eu tendance à interrompre leurs carrières davantage que leurs collègues masculins, par exemple. On notera d'ailleurs que, pour l'entreprise en question, l'utilisation irraisonnée d'un tel algorithme aurait pour conséquence de se priver de talents. Le problème éthique croiserait ici directement l'enjeu d'efficacité.

> Inconscients chez ceux-là mêmes qui sélectionnent les données, les biais ne sont pas forcément sensibles pour les utilisateurs qui y sont sujets

Dès lors, l'opération même d'entraînement des algorithmes – à travers la sélection qu'elle suppose des données à prendre en compte – apparaît comme le cœur d'un enjeu éthique et juridique, et non pas seulement technique ou d'efficacité. Cet enjeu recoupe en partie celui de la délégation de prises de décisions, abordé précédemment : choisir quelles données sont utilisées pour les phases d'apprentissage revient bien à prendre des décisions parfois lourdes de conséquences. En revanche, le caractère spécifique de l'enjeu abordé ici tient au fait qu'il **s'agit de décisions et de choix qui peuvent être effectués de manière presque inconsciente** (alors que le codage d'un algorithme classique

### LE SAVIEZ-VOUS ?

**À l'occasion du débat organisé le 24 juin 2017 par le Génotoul (Toulouse), Philippe Besse, Professeur de mathématiques et de statistique à l'Université de Toulouse a souligné que nous ne sommes pas tous égaux devant la médecine personnalisée, car les bases de données utilisées à l'heure actuelle sont largement biaisées : une étude a révélée qu'en 2009, 96 % des échantillons de ces bases ont des ancêtres européens (la démonstration porte sur 1,5 million d'échantillons). D'autres sources de biais sont l'âge (car toutes ces bases de données sont largement occupées par des personnes relativement âgées) et le genre, plusieurs publications récentes insistant sur l'importance de l'effet du genre sur le développement des maladies concernées. Dans ces bases, le chromosome X est largement sous représenté et le Y est quasiment absent. Philippe Besse conclut ainsi : « si vous êtes une femme d'origine africaine et jeune, je ne pense pas que la médecine personnalisée vous concerne ».**

21 https://www.bloomberg.com/graphics/2016-amazon-same-day/

et déterministe est toujours une opération délibérée). Celui qui entraîne un algorithme y insère d'une certaine façon sa propre vision du monde, ses valeurs ou, à tout le moins, des valeurs présentes plus ou moins directement dans les données tirées du passé. La chercheuse Kate Crawford, notamment, a ainsi mis en évidence l'endogamie sociale, raciale et de genre qui caractérise les milieux où se recrutent ceux qui entraînent aujourd'hui l'intelligence artificielle[22].

Tout ceci explique largement l'une des caractéristiques les plus problématiques de ces biais et des discriminations auxquelles ceux-ci peuvent donner lieu : ils sont souvent particulièrement difficiles à découvrir. Inconscients chez ceux-là mêmes qui sélectionnent les données, ils ne sont pas forcément sensibles pour les utilisateurs qui y sont sujets. Le caractère ciblé des offres d'emploi évoquées précédemment fait que les femmes concernées n'avaient pas connaissance des offres d'emploi proposées aux hommes. C'est l'une des conséquences du phénomène d'« enfermement algorithmique », dont il sera question plus loin. Enfin, les systèmes d'intelligence artificielle font quant à eux des choix dont la logique (voire l'existence même) échappe à leurs concepteurs.

En somme, les biais et les discriminations générées par les algorithmes soulèvent aujourd'hui deux questions majeures. Faut-il d'abord considérer au moins dans certains cas que l'intelligence artificielle ne fait jamais que reconduire des biais et des discriminations déjà existants dans la société ? En d'autres termes, les algorithmes ne seraient jamais ici que des « conducteurs » de biais, ils ne feraient que les répéter sans les créer eux-mêmes. On pourrait à tout le moins objecter à une telle position que l'échelle à laquelle ils se déploient et leur impact potentiel en font les lieux privilégiés pour la lutte contre les discriminations, que leur puissance, en somme, implique des obligations renforcées. Sans compter qu'il n'est pas exclu qu'ils puissent aussi avoir un effet démultiplicateur de ces biais.

Deuxièmement, **comment se donner les moyens de repérer effectivement ces biais, dont nous avons souligné le caractère parfois invisible ?** Faut-il d'ailleurs distinguer entre des biais qui seraient acceptables et d'autres que la société ne pourrait pas tolérer (comme ceux évoqués plus haut) ? Enfin, comment lutter efficacement contre ces biais et s'assurer que les algorithmes respectent les valeurs fondamentales élaborées démocratiquement par nos sociétés ?

Il faut enfin souligner ici une dimension que nous verrons resurgir dans la suite de ce rapport : les impacts non pas seulement individuels (sur la personne), mais également collectifs que peuvent avoir les algorithmes. L'exemple de l'exclusion par un service d'Amazon de quartiers entiers en offre une illustration.

# Fragmentation algorithmique : la personnalisation contre les logiques collectives

L'omniprésence des algorithmes, notamment ceux liés à notre navigation sur le Web et sur les réseaux sociaux, est indissociablement liée à la dynamique de personnalisation des contenus et des services. Cette personnalisation au service de l'individu recèle cependant une dimension problématique en portant potentiellement atteinte à des logiques proprement collectives sur lesquelles reposent nos sociétés, de la structuration de l'espace public démocratique aux mécanismes de mutualisation dans l'ordre économique. Alors que l'impact des algorithmes sur les personnes est un phénomène bien repéré et pris en compte par la loi depuis longtemps, ses impacts collectifs posent également question aujourd'hui.

## Enfermement algorithmique et perte de pluralisme culturel

Le thème de l'enfermement algorithmique a fait l'objet de nombreuses discussions depuis l'ouvrage d'Eli Pariser sur la « bulle filtrante[23] ». Il renvoie à l'idée selon laquelle l'activité indispensable jouée par les algorithmes en termes

22 Kate Crawford, "Artificial Intelligence's White Guy Problem", *The New York Times*, 25 juin 2016.
23 Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, New York, Penguin Press, 2011.

de classement et de filtrage d'une information devenue surabondante aurait pour effet indirect de nuire au pluralisme et à la diversité culturelle: en filtrant les informations, en s'appuyant sur les caractéristiques de leurs profils, **les algorithmes augmenteraient 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.**

Le thème de la bulle filtrante se pose à deux échelles, celle des individus et celle de la société dans son ensemble.

**À l'échelle de l'individu,** le risque est que celui-ci se voie purement et simplement assimilé à un alter ego numérique constitué à partir de ses données et se trouve en quelque sorte enfermé dans une bulle de recommandations toujours conforme à ce profil. Les effets d'une offre culturelle et de contenus plus abondante que jamais auparavant se verraient ainsi paradoxalement neutralisés par un phénomène de limitation de l'exposition effective des individus à la diversité culturelle. Un tel phénomène pourrait d'ailleurs se produire alors même que l'individu souhaiterait en principe une telle diversité. La Direction Générale des Médias et des Industries Culturelles (DGMIC) souligne ainsi que « la recommandation algorithmique est fondée sur la consommation réelle des utilisateurs plutôt que sur leurs désirs ou aspirations ».

Il faut pourtant relever que d'importants spécialistes, chercheurs et praticiens du numérique contestent l'idée d'enfermement algorithmique ou du moins invitent à poser la question de manière plus nuancée. Ainsi, selon Antoinette Rouvroy, « 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 » (propos tenus le 23 janvier 2017 lors de l'événement de lancement du débat public à la CNIL). Il est vrai que l'on constate que cette potentialité n'est de fait guère exploitée. En effet, la consommation culturelle repose sur une structure duale de goûts : d'une part des liens forts « traduisant une préférence avérée pour un type de contenus bien identifié a priori », d'autre part des liens faibles « rendant compte d'une affinité non encore révélée pour un type de contenus restant à découvrir à posteriori[24] ». Or, la plupart des algorithmes prédictifs des grandes plateformes culturelles (Netflix, Amazon, Spotify, etc.) se focalisent sur les liens forts. Aucune des grandes catégories d'algorithmes n'envisage la sérendipité comme variable essentielle aux choix de consommation.

> Les algorithmes augmenteraient 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

Dominique Cardon souligne quant à lui que « le numérique a apporté une diversité informationnelle 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. [...] Un risque existe que se produisent 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 » (propos tenus le 23 janvier 2017 lors de l'événement de lancement du débat public à la CNIL). Enfin, la DGMIC estime que les incitations concurrentielles à la différenciation ainsi qu'*« une vision libérale de l'individu considérant l'étendue du choix comme un facteur d'épanouissement »*[25] pourraient limiter les risques pesant sur la diversité en incitant les acteurs à se saisir de l'enjeu de l'enfermement et à lui apporter des réponses.

**À l'échelle de sociétés** considérées dans leur ensemble, **les formes de privation d'exposition des individus à l'altérité, à des opinions différentes des leurs, notamment dans le registre politique, pourraient en tout cas constituer, selon certains, un problème pour la qualité et la vitalité du débat public, pour la qualité et la diversité de l'information, terreaux du fonctionnement correct des démocraties.**

À l'horizon logique du phénomène, la personnalisation de l'information aurait pour conséquence une fragmentation extrême de l'espace public, la disparition d'un socle minimum d'informations partagées par l'ensemble du corps politique et permettant la constitution d'un véritable débat.

**24** Rapport du CSA Lab
**25** Natali HELBERGER, Kari KARPPINEN & Lucia D'ACUNTO, "Exposure diversity as a design principle for recommender systems", Information, Communication & Society, 2016.

À l'heure où une part croissante des citoyens utilisent les réseaux sociaux comme le principal (et parfois seul) moyen d'information[26], l'enjeu est important pour la pérennité de la vie démocratique. Si la tendance à s'entourer de personnes partageant les mêmes idées et les mêmes valeurs n'est pas nouvelle, du moins la presse traditionnelle avec sa logique éditoriale permet-elle au lecteur d'avoir une plus claire conscience de l'orientation du contenu qu'il consomme. Les débats portant sur ce sujet font pourtant clairement ressortir que les effets dénoncés sous la rubrique de la « bulle de filtre » ne sont pas fatalement et toujours produits par les algorithmes. Ils sont avant tout le résultat du paramétrage d'algorithmes que l'on pourrait tout aussi bien programmer autrement et à qui l'on pourrait, à l'inverse, donner comme objectif d'exposer les individus à une diversité culturelle, informationnelle, politique forte.

Il est possible que la nature même du problème en ait ralenti la prise de conscience publique. À la limite, en effet, l'individu peut très bien vivre dans sa bulle informationnelle sans en prendre conscience. Le confort provoqué par l'absence de contradiction ou encore le biais de confirmation caractérisant l'esprit humain et que connaissent bien les sciences cognitives ne sont évidemment pas des facteurs propices à la remise en cause de l'enfermement algorithmique. Autrement dit, rien ne prédispose l'individu à s'apercevoir qu'il est pris dans une bulle informationnelle. Il n'est dès lors guère étonnant que les mises en cause de ce phénomène s'accompagnent souvent de récits relatant le moment de sa prise de conscience, un moment s'apparentant à un choc. C'est ainsi que les débats sur la bulle filtrante et ses effets politiques ont été notamment relancés à l'occasion de la campagne présidentielle américaine de 2016 ainsi que par celle du Brexit, quelques mois avant. Deux chocs électoraux à l'occasion desquels de nombreux internautes partisans d'Hillary Clinton ou opposants au Brexit ont été particulièrement frappés de constater des résultats que leurs fils d'actualité ne laissaient en rien présager. Plus récemment, en août 2017, la sociologue Zeynep Tufekci, spécialiste des mouvements de contestation en ligne a remarqué – parmi d'autres – que son fil d'information Facebook demeurait silencieux sur les événements de Ferguson au moment même où elle voyait le hashtag Ferguson se répandre sur Twitter.

On peut considérer que **l'absence de compréhension claire par les individus du fonctionnement des plateformes qu'ils utilisent pour s'informer, notamment, fait partie intégrante du problème.** Une étude a ainsi montré que plus de 60% des utilisateurs de Facebook n'ont aucune idée de l'activité éditoriale que joue effectivement l'algorithme et croient que tous les posts de leurs amis et des pages qu'ils suivent apparaissent sur leur fil d'actualités[27]. En vérité, ils n'en voient que 20%, sélectionnés selon plusieurs facteurs : promotion publicitaire du post, interactions passées de l'utilisateur avec des posts considérés comme similaires – *like,* commentaire, partage-, nombre d'autres utilisateurs ayant fait de même, etc.

L'usage fait des algorithmes par l'économie numérique à des fins de personnalisation du service et de l'expérience répond donc à une logique qui pose problème dès lors que l'on considère ses effets d'un point de vue, non plus seulement économique, mais aussi culturel ou politique. **L'objet des grandes plateformes algorithmiques est la satisfaction d'un consommateur, d'un *homo economicus*. Les effets politiques et culturels à grande échelle de leurs algorithmes ne leur posent question que secondairement.**

## Atomisation de la communauté politique

Cet effet induit des algorithmes et de leur fonction de personnalisation peut néanmoins devenir un levier direct pour certains acteurs qui cherchent à les exploiter à des fins d'influence, voire de manipulation. Les *fake news,* largement évoquées lors de la campagne menée par Donald Trump, si elles ne sont pas un produit direct des algorithmes, se diffusent et s'amplifient à l'intérieur des chambres d'écho constituées par les algorithmes des réseaux sociaux ou des moteurs de recherche. Plus directement encore, des logiciels de stratégie politique de plus en plus élaborés et appuyés sur un ciblage de plus en plus fin des électeurs conduisent à une fragmentation potentiellement sans précédent d'un discours politique adressé désormais à des individus atomisés. Les pratiques de la société Cambridge Analytica, qui a travaillé pour le candidat Trump, représentent la pointe de diamant de ces nouveaux usages des algorithmes à des fins électorales (voir encadré). La tendance à la fragmentation personnalisée du discours politique, appuyée sur la capacité croissante de l'IA à composer des messages en fonction des différents profils, pose aujourd'hui de sérieuses questions. Faut-il y voir une forme de manipulation ? Faut-il y poser des limites ? Faut-il considérer ces pratiques comme le fruit inéluctable et difficilement régulable de l'évolution technologique et dès lors imaginer des contrepoids ? Si oui, lesquels ?

On le voit, le thème de l'enfermement est l'envers de celui de la personnalisation algorithmique. Ceci explique qu'enfermement et fragmentation puissent être aussi décelés dans des secteurs autres que celui de la consommation culturelle et des médias ou de la politique.

**26** Selon le Pew Research Center, 61% des "millenials" utilisent Facebook comme leur première source d'information sur la politique l'action gouvernementale (Pew Research Center, *Millenials & Political News. Social Media – the Local TV for the Next Generation ?,* juin 2015).
**27** http://www-personal.umich.edu/~csandvig/research/Eslami_Algorithms_CHI15.pdf

FOCUS

## Algorithmes et stratégie électorale

Les dernières élections présidentielles, aux États-Unis mais aussi en France, ont donné lieu à l'utilisation croissante des **logiciels de stratégie électorale** reposant sur la mise en œuvre d'algorithmes prédictifs d'analyse des données électorales. Loin des méthodes plus traditionnelles de campagne, des messages politiques très ciblés peuvent désormais être adressés aux électeurs. C'est aux Etats-Unis que l'on peut identifier les exemples les plus accomplis d'un tel profilage individuel. Dès les élections présidentielles de 2008 et 2012, les équipes électorales de Barack Obama disposaient de centaines de données sur la quasi-totalité des électeurs. En 2016, grâce à l'analyse des données issues des réseaux sociaux et des courtiers en données, Cambridge Analytica aurait pu envoyer pour le compte du candidat Trump des milliers de messages extrêmement individualisés au cours d'une même soirée[28]. Si cette entreprise a par la suite tenu un discours tendant à minimiser ses premières affirmations, cette affaire n'en est pas moins révélatrice d'une tendance de fond susceptible de s'approfondir à l'avenir.

En France, les **principes de protection des données à caractère personnel** limitent toutefois dans les faits le développement de tels logiciels de ciblage individuel, le consentement constituant un prérequis essentiel à une telle collecte. La CNIL a d'ailleurs rappelé, dans un communiqué de novembre 2016, les règles pour l'utilisation des données issues des réseaux sociaux à des fins de communication politique[29].

## L'enfermement algorithmique, un enjeu transversal

La question de l'enfermement algorithmique ne se limite pas aux secteurs de la culture, de l'information ou de la politique. En effet, l'intrication des fonctions de prédiction et de recommandation présentes dans les usages des systèmes algorithmiques aujourd'hui modelés par l'écosystème numérique est susceptible de générer des prophéties auto-réalisatrices pouvant enfermer les individus dans un destin « prédit ».

Une forme d'enfermement n'est-elle pas une conséquence possible de futurs usages des *learning analytics* et de l'*adaptative learning* (ou éducation personnalisée) ? Sans remettre en cause les promesses de ces techniques, il est légitime de s'interroger sur les effets que pourraient avoir des systèmes prétendant définir des parcours d'apprentissage sur la base du profil de chaque élève et de la prédiction élaborée à partir de l'application d'un modèle mathématique à ce profil. N'y a-t-il pas un risque que la prédiction devienne auto-réalisatrice et que l'élève se trouve assigné à un destin scolaire et professionnel dès lors que le diagnostic aura été posé ? Comme le souligne Roger-François Gauthier, « avec les *learning analytics,* la prédiction pourrait déboucher sur un enfermement des élèves. En France, 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[30] ».

On peut, de la même façon, rattacher à l'idée d'enfermement algorithmique certains impacts possibles de l'utilisation des algorithmes dans le secteur des ressources humaines et du recrutement. Laurence Devillers évoque ainsi le risque de « normalisation des profils » que pourrait faire courir l'algorithme, du moins un usage non raisonné de l'algorithme, au recruteur. C'est en quelque sorte ce dernier qui serait victime ici d'enfermement dans des profils prédéfinis à l'avance, se privant de la part de sérendipité inhérente au processus de recrutement dans la mesure où celui-ci peut permettre de repérer des profils atypiques, non conformes aux critères définis *a priori,* mais finalement intéressants. Comment repérer de tels profils si une part croissante de la sélection des candidats se trouve déléguée à des systèmes automatiques ?

28 https://www.theguardian.com/politics/2017/feb/26/robert-mercer-breitbart-war-on-media-steve-bannon-donald-trump-nigel-farage
29 https://www.cnil.fr/fr/communication-politique-quelles-sont-les-regles-pour-lutilisation-des-donnees-issues-des-reseaux
30 Propos tenus à l'occasion du lancement du débat public, le 23 janvier 2017, à la CNIL.

## Démutualisation

La personnalisation algorithmique soulève un enjeu spécifique au secteur de l'assurance. En effet, **la dynamique de personnalisation des offres et des services ne conduit-elle pas à une remise en cause de la mutualisation, c'est-à-dire de la logique même de l'assurance et du pacte social sur lequel elle repose ?** Que plusieurs individus acceptent de s'assurer, c'est-à-dire de mettre en commun leurs risques, suppose que ces risques leur demeurent au moins partiellement opaques. Je m'assure en ignorant lequel de moi ou de mon voisin contractera une maladie occasionnant de lourds frais de santé. La segmentation accrue que rendrait possible l'utilisation des masses de données générées par les comportements des individus en ligne (réseaux sociaux, notamment) ou hors-ligne (données issues de bracelets connectés, par exemple) tendrait à lever le « voile d'ignorance[31] » sous-tendant la mutualisation assurantielle et que contribue à maintenir une segmentation sommaire.

Ces innovations ne déboucheront-elles pas sur de nouvelles formes de discrimination et d'exclusion ? Les individus jugés « à risque » pourraient se voir appliquer des tarifs plus élevés, voire même être victimes de décisions de refus d'assurance. À cela s'ajoute le fait que l'établissement d'une corrélation entre un comportement et le risque de survenue d'une pathologie pourrait aboutir à défavoriser les individus ayant des comportements jugés « à risque » (consommation de tabac, nourriture jugée trop grasse, trop sucrée, etc.). La question serait alors celle des limites à poser à ce qui peut apparaître comme une normalisation excessive des comportements des personnes lorsque ceux-ci seraient estimés « mauvais ». Les algorithmes, via les corrélations qu'ils établissent dans les données, finiraient par édicter la norme des comportements individuels, une norme à laquelle on ne pourrait échapper qu'au prix d'un renchérissement de l'assurance. À la différence d'un mécanisme comme l'augmentation des prix du tabac (dont la consommation est considérée comme un coût pour la collectivité), de tels arbitrages échapperaient à la délibération collective et surgiraient des données mêmes. Par ailleurs, une telle approche évacuerait complètement les déterminants collectifs et sociaux des comportements pour ne plus mettre en exergue que la seule responsabilité des individus. Quant à d'autres facteurs de risque, liés à l'environnement de l'individu ou à son patrimoine génétique, ils seraient susceptibles de déboucher sur une discrimination et une exclusion inévitables dans la mesure où les personnes concernées n'auraient aucune prise sur eux.

Si la course aux « bons risques » pourrait donc être accrue entre les assureurs, il est cependant douteux que celle-ci soit favorable à ces derniers pris dans leur ensemble. L'assureur aurait intérêt à la mutualisation. Selon Florence Picard, de l'Institut des Actuaires, *« plus il segmente fermement les groupes, plus il prend le risque de mettre fin à la mutualisation. Son but est que le risque soit maîtrisable: plus on segmente, plus on prend le risque de se tromper[32] ».*

# Entre limitation des mégafichiers et développement de l'intelligence artificielle : un équilibre à réinventer

Le fonctionnement des algorithmes auxquels nous avons quotidiennement recours repose sur le traitement de nombreuses données, dont une grande part de données personnelles, traces numériques laissées par nos navigations en ligne, par l'utilisation de nos smartphones, de nos cartes de crédit, etc. **La recherche d'une performance accrue des algorithmes est un facteur allant dans le sens d'une collecte croissante, d'un traitement et d'une conservation accrus de données à caractère personnel.**

On peut ainsi se demander si le développement de l'intelligence artificielle n'est pas susceptible, à un certain stade, d'entrer en tension avec les principes éthiques inscrits dans la législation depuis la loi Informatique et libertés. L'intelligence artificielle est grande consommatrice de données ; elle a besoin d'une grande mémoire (autrement dit, de bases de données qu'elle va conserver sur une période aussi longue que possible). Les principes de la loi de 1978 renvoient, quant à eux, par le truchement du principe de finalité, à une minimisation de la collecte de données per-

31 Antoinette Rouvroy déplace ainsi, en l'appliquant au domaine de l'assurance, le concept forgé par John Rawls pour établir une expérience de pensée destinée à envisager un problème moral.

32 « Algorithmes et risques de discriminations dans le secteur de l'assurance », manifestation organisée par la Ligue des Droits de l'Homme le 15 septembre 2017.

sonnelles ainsi qu'à la limitation de la durée de conservation de ces données comme à des garanties nécessaires à la protection des personnes et de leurs libertés.

Certes, les principes de la loi de 1978 (repris dans le Règlement général sur la protection des données, qui entrera en application en mai 2018) constituent un équilibre général, offrant une certaine souplesse à l'ensemble. Des mesures de sécurité renforcées peuvent dans une certaine mesure être considérées comme un contrepoids à une durée de conservation allongée des données. Il n'est pourtant pas certain que l'ampleur des transformations technologiques induites par le développement de l'intelligence artificielle ne remette pas en cause ce schéma.

Par exemple, la médecine de précision semble lier ses progrès à la constitution de bases de données toujours plus larges, à la fois en termes de nombres d'individus concernés qu'en termes de nombre et de variété de données conservées sur chacun d'entre eux. L'épigénétique prétend ainsi croiser une approche par les données génétiques de l'individu à une approche prenant en compte les données environnementales, celles concernant le milieu, voire le mode de vie du « patient » (si tant est que cette notion ait encore un sens dans un contexte de plus en plus orienté vers la « prédiction »). La médecine de précision repose sur l'idée de profiler le plus finement possible ce dernier et la pathologie dont il est affecté afin de comparer ce profil à ceux d'autres individus au profil très proche, de façon à identifier le traitement le plus approprié à ce patient. À la limite, on pourrait aller jusqu'à considérer que l'objectif sanitaire poursuivi implique la constitution d'immenses bases de données. Or, rien n'indique où devrait s'arrêter la collecte de données : au dossier médical ? Au génome ? Aux données épigénétiques, c'est-à-dire environnementales (habitudes de vie, habitat, alimentation, etc.) ? En remontant à combien d'années ? Notons que ce type de problème n'est nullement propre à la médecine. Il se poserait sous un aspect proche dans le domaine de la sécurité, où l'impératif de repérage des suspects semble justifier une collecte de données toujours plus massives sur les individus.

On voit bien que **la question posée ici est celle de l'équilibre à trouver entre protection des libertés (protection des données personnelles) et progrès médicaux.** Il ne saurait être question d'y répondre ici, tant elle mériterait de faire l'objet d'une réflexion poussée. Celle-ci devrait d'ailleurs nécessairement impliquer une évaluation des progrès effectivement à attendre de la médecine de précision. Ainsi Philippe Besse, professeur de mathématiques à l'Université de Toulouse, considère que les données mises à la disposition de la recherche médicale dans le cadre du Système National des Données de Santé (SNDS) sont suffisantes pour accomplir des progrès que la complexité du vivant limitera de toute façon bien en-deçà de ce qu'annoncent certaines prophéties[33].

# Qualité, quantité, pertinence : l'enjeu des données fournies à l'IA

Les systèmes algorithmiques et l'intelligence artificielle reposent sur l'utilisation de données (personnelles ou non) qui leur sont fournies en entrée et qu'ils traitent pour produire un résultat. Schématiquement, cette caractéristique soulève trois enjeux connexes mais distincts : celui de la qualité, celui de la quantité et celui de la pertinence des données fournies à ces systèmes.

La **question de la qualité des données utilisées par les algorithmes et l'IA** est la plus simple. Il est facile de comprendre que **des données erronées ou tout simplement périmées impliqueront en bout de chaîne des erreurs ou des dysfonctionnements plus ou moins graves selon le domaine concerné,** du simple envoi de publicités ciblées correspondant mal à mon profil réel jusqu'à une erreur de diagnostic médical. Assurer la qualité de la donnée entrante dans les systèmes algorithmiques et d'intelligence artificielle constitue donc un enjeu appelé à prendre une importance de plus en plus cruciale au fur et à mesure que ces machines vont être amenées à prendre une autonomie croissante. Or, assurer la qualité de la donnée est coûteux. La corruption des données peut être le résultat aussi bien d'un problème technique très matériel impliquant l'état des capteurs affectés à leurs collecte que d'un problème humain lié à l'intérêt de certains acteurs à biaiser les données qu'ils sont chargés d'entrer dans le système. La tentation de la négligence à cet égard doit être prise au sérieux, notamment dans des domaines où l'impact de données de

**33** Débat organisé par le Génotoul à Toulouse, le 24 juin 2017.

mauvaise qualité pourrait n'être pas immédiatement sensible, comme dans le secteur du recrutement, par exemple. Les données des réseaux sociaux professionnels, parfois considérées comme une manne inépuisable, posent à cet égard des problèmes de fiabilité (liés à la tendance des individus à embellir leur CV ou au contraire à des absences de mise à jour). La confiance accordée par l'utilisateur au résultat produit par une machine jugée objective et plus performante que l'homme est un facteur supplémentaire pouvant favoriser la négligence.

La **quantité de données disponibles** peut constituer un autre facteur néfaste à la qualité des résultats fournis par les systèmes algorithmiques et d'intelligence artificielle. Cathy O'Neil évoque ainsi l'exemple d'une collectivité ayant recouru aux États-Unis à un logiciel d'évaluation des enseignants. L'utilisation de ce logiciel s'est notamment soldée par le licenciement d'enseignants dont la qualité était pourtant de notoriété publique dans les communautés locales au sein desquelles ils évoluaient. L'une des raisons essentielles en est que l'algorithme utilisé pour évaluer la progression annuelle des élèves de chaque enseignant aurait besoin de bien plus que des données concernant tout au plus quelques dizaines d'élèves. Dans un cas où les variables susceptibles d'expliquer, à côté de la performance du professeur, les mauvais résultats d'un élève (difficultés relationnelles, problèmes familiaux, problèmes de santé, etc) sont si nombreuses, un nombre si limité de cas ne peut avoir aucune valeur statistique. La seule valeur de ce résultat est de donner le sentiment aux décideurs de prendre des décisions rationnelles, objectives et efficaces car s'autorisant du prestige de la machine.

Cela ne signifie toutefois nullement que l'accumulation irréfléchie de données doive constituer un objectif en soi. Dans certains cas, en effet, la variété des données sera plus précieuse que leur simple quantité. Par exemple, les données de millions de véhicules suivant la même route seront moins utiles à l'algorithme d'une application GPS que des données en bien moins grand nombre de véhicules empruntant des itinéraires plus variés.

Enfin, **la question de la pertinence des données renvoie moins à la véracité de ces dernières qu'aux biais qui peuvent présider à leur collecte.** Comme cela a été montré précédemment (Voir « Biais, discriminations et exclusion »), il peut être tout à fait exact que très peu de femmes aient mené à bien une carrière de haut niveau dans telle ou telle entreprise. En revanche, prendre ce résultat comme indicatif de la capacité de femmes à accomplir à l'avenir de brillantes carrières dans cette même entreprise relève bien évidemment d'une approche biaisée. En l'occurrence, le jeu de données envisagé ici intègre des formes d'inégalités et/ou de discriminations. Ignorer ce type de biais reviendrait à perpétuer ou à laisser se perpétuer ces phénomènes.

On voit à travers ces trois enjeux que les promesses des algorithmes ne peuvent être tenues qu'au prix d'une grande rigueur dans la collecte et le traitement des données utilisées. Qu'une telle exigence de rigueur (et d'investissement matériel et humain) puisse ne pas être respectée par certains acteurs représente un risque évident, alors même que les algorithmes sont souvent présentés comme sources d'une vérité « objective », « neutre ». Dans l'exemple de l'algorithme utilisé pour évaluer les professeurs aux États-Unis évoqué par Cathy O'Neil, **la négligence méthodologique des concepteurs et promoteurs de l'algorithme a pour corollaire la confiance exagérée, dénuée d'esprit critique** qu'accordent à ce dernier des utilisateurs dont l'attention se focalise sur la seule nécessité d'obtenir un quota de professeurs à éliminer du système. Pourtant, si assurer la qualité et la pertinence des données fournies aux algorithmes s'impose donc comme une exigence éthique, cette dernière constitue bien à terme une condition de l'utilité durable des algorithmes pour leurs utilisateurs et pour la société en général.

ENQUÊTE

## La crainte devant les risques des algorithmes et de l'IA augmente avec l'âge*

**Les jeunes sont plus sensibles aux opportunités** portées par l'algorithme : 68,5 % des 18-24 ans considèrent que les opportunités surpassent les potentielles menaces. En revanche, seul 36 % des 55-64 ans estiment que les bénéfices sont plus importants que les risques.
Certaines applications des algorithmes sont mieux acceptées chez les plus jeunes: 75 % des 18-24 ans regardent favorablement des recommandations en vue d'achats en ligne (contre 48 % pour l'ensemble du panel), 50 % en vue du choix de l'âme-sœur (contre 26 %).

* Enquête réalisée dans le cadre du débat public par l'association « Familles rurales », association familiale orientée vers les milieux ruraux, auprès de 1076 de ses adhérents.

# L'identité humaine au défi de l'intelligence artificielle

L'autonomisation des machines, d'une part, l'hybridation croissante des humains avec la machine, d'autre part, questionnent l'idée d'une spécificité humaine irréductible.

## Des machines éthiques ?

La première zone de porosité entre humains et machines s'établit autour de la question de l'idée de machine éthique. En effet, une façon radicale d'aborder les questions soulevées par l'éventuelle délégation de décisions à des machines autonomes (intelligence artificielle) est d'envisager que de rendre les machines « éthiques » serait une solution aux problèmes évoqués plus haut dans ce rapport. Une telle piste de réflexion est liée à la question de savoir s'il est même possible de formaliser une éthique[34] afin de la programmer dans une machine. Autrement dit, **peut-on automatiser l'éthique ?** Ce problème est apparu au cours des débats comme l'un de ceux retenant particulièrement l'attention de la communauté des chercheurs en intelligence artificielle. Gilles Dowek (CERNA) l'a ainsi souligné lors de la journée d'étude organisée au Collège des Bernardins le 20 septembre 2017.

Le fameux dilemme du tramway est très souvent évoqué à l'occasion de réflexions portant sur ce problème. On sait que ce dilemme met en scène un tramway sans freins dévalant une pente ; le tramway arrive devant un embranchement ; selon qu'il s'engage sur l'une ou l'autre des deux voies, il tuera une personne ou bien plusieurs. Dès lors, quelle devrait être la conduite d'une personne ayant la possibilité de manœuvrer l'aiguillage et donc de choisir, pour ainsi dire, l'un des deux scénarios possibles ? L'intérêt de cette expérience de pensée est qu'elle peut donner lieu à toute une gamme de variations : qu'en est-il si la personne seule attachée à l'une des deux voies se trouve être un proche parent ? Si les personnes sur l'autre voie se trouvent être 5 ou bien 100 ?

On voit aisément comment ce dilemme peut être adapté à l'hypothèse de voitures autonomes qui seraient mises en circulation prochainement : selon quels principes une voiture placée dans une situation de dilemme éthique de ce type devrait-elle « choisir » de se comporter ? Le dilemme du tramway a l'intérêt de mettre en évidence le fait que différents choix « éthiques » sont possibles. Dès lors que des dilemmes de ce type auraient été anticipés au stade du développement du système, il serait bien sûr possible de leur donner une réponse. Mais précisément, **la spécificité de l'éthique n'est-elle pas de concerner des situations inédites, impliquant éventuellement des conflits de valeurs dont la solution doit être élaborée par le sujet** (pensons à Antigone, prise entre éthique familiale et éthique civique) ? N'est-elle pas de s'élaborer toujours en situation ? Dès lors l'hypothèse d'une formalisation de l'éthique n'est-elle pas quelque peu illusoire ? À tout le moins, elle implique une conception implicite de l'homme qui n'a rien d'évident.

Retenons du moins que, pour l'heure, des expressions comme « éthique des algorithmes » ou « algorithmes éthiques » ne doivent pas être prises au pied de la lettre et comportent une part d'anthropomorphisme revenant à attribuer des capacités humaines à des machines. Certains considèrent qu'elles sont susceptibles de fausser un débat qui devrait se concentrer sur les exigences à l'égard des hommes qui conçoivent, entraînent, déploient et utilisent les systèmes algorithmiques et d'intelligence artificielle.

Elles ne constitueraient alors qu'une métaphore commode mais à ne pas entendre littéralement. À l'inverse, comme le rappelle par exemple Gilles Dowek, on peut considérer comme légitime le recours à ce type de métaphores dans la mesure où elles reviennent à prendre acte de l'autonomie croissante de ces systèmes et de la nécessité de formaliser, autant que faire se peut, une éthique et de la programmer dans des algorithmes. Quoi qu'il en soit, même si une éthique en tant que telle pouvait être encodée dans une machine (c'est-à-dire si cette dernière avait la possibilité de ne pas seulement répondre d'une certaine façon à une situation éthique envisagée à l'avance lors de son développement mais bien d'aborder des situations nouvelles en leur appliquant un raisonnement éthique), le choix du type d'éthique à encoder resterait bien, en dernière analyse, du ressort de l'homme. Le vrai enjeu est alors de s'assurer que les choix éthiques faits au stade du développement ne font pas l'objet d'une confiscation par « une petite caste de scribes » (Antoine Garapon). L'échelle de déploiement des algorithmes à l'heure du numérique en fait une question démocratique essentielle.

**34** C'est-à-dire une règle générale d'évaluation de la conduite à adopter face à toute situation – éthique déontique ou éthique conséquentialiste – ou un corpus de règles remplissant la même fonction – éthique kantienne, éthique bouddhiste, etc.

## L'hybridation de l'homme et de la machine : repenser l'identité humaine ?

Une façon d'envisager la question éthique appliquée aux algorithmes et à l'intelligence artificielle peut être de confronter ces derniers à l'affirmation – présente à l'article premier de la loi Informatique et libertés – selon laquelle l'informatique « ne doit pas porter atteinte à l'identité humaine ».

Les pages précédentes ont abordé des problèmes liés à la façon dont l'homme agence son action avec des artefacts, question ancienne mais renouvelée par l'émergence d'artefacts dotés d'une « autonomie » croissante à l'heure des algorithmes et de l'intelligence artificielle[35]. Ces propos soulignent en effet que le développement de ces technologies, selon la manière dont il s'opérera, peut affecter l'une des composantes de l'identité et de la dignité humaines, à savoir sa liberté et sa responsabilité. La montée en puissance d'une forme d' « autonomie » machinique doit bien sûr être fortement nuancée. Gérard Berry, professeur au Collège de France et titulaire de la chaire « Algorithmes, machines et langages » rappelle ainsi : « un jour, nous dit-on, les machines parleront et seront autonomes, le numérique donnera naissance à une nouvelle forme de vie. La date pour l'autonomie des machines et leur capacité de parole créative, personne ne la donne, et je ne la connais pas, loin de là. Surtout, de quelle vie parlons-nous ?[36] ». Néanmoins, on pourrait se demander si la trajectoire technologique d'ores et déjà à l'œuvre ne devra pas conduire à questionner la pertinence de la notion même d' « identité humaine », dans la mesure où celle-ci implique une séparation étanche entre humain et non-humain. La question du « droit des robots » d'ores et déjà soulevée par des juristes et récemment examinée par le Parlement européen (rapport Delvaux) a pour horizon ce brouillage possible des frontières de l'humain. À de tels arguments post-humanistes, la tradition humaniste pourrait certes rétorquer que l'autonomie machinique n'est aujourd'hui qu'un leurre, une métaphore destinée à styliser un objet complexe et masquant finalement une responsabilité et une action humaines certes diluées, éclatées, mais bien réelles.

> Le développement de ces technologies peut affecter l'une des composantes de l'identité et de la dignité humaines, à savoir sa liberté et sa responsabilité

Si une première hybridation entre l'homme et la machine s'opère au plan de l'action, la réflexion devra aussi nécessairement s'élargir à l'avenir pour prendre en compte l'hybridation physique parfois annoncée entre algorithmes, humains, voire animaux (avec l'adjonction d'implants intelligents et communicants). Cette hybridation physique est une étape supplémentaire de l'évolution déjà à l'œuvre dans l'interaction permanente qui nous lie d'ores et déjà à une foule de processus algorithmiques.

Enfin, ce thème d'une subversion éventuelle de la frontière entre l'homme et les choses (ou plutôt, entre l'homme et la machine) trouve déjà une réalité extrêmement concrète au plan phénoménologique dans certaines tentatives récentes d'applications de la robotique qui s'illustrent d'abord dans la forme humaine donnée aux robots. On pense ici au robot Pepper de la firme Aldebaran, destiné à être déployé dans des espaces commerciaux pour interagir avec les clients. Surtout, et ceci concerne directement le sujet des algorithmes et de l'intelligence artificielle, **tout un champ de recherche vise à créer des robots empathiques capables de percevoir les émotions des humains** (par l'analyse du visage, de la voix, etc.) de façon à s'adapter à leur interlocuteur. La première question posée par ces recherches est évidemment celle de la limite entre, d'une part, les apports bénéfiques d'une intelligence artificielle capable de comprendre et de s'adapter aux états émotionnels de ses interlocuteurs et, d'autre part, une forme de manipulation appuyée sur une ingénierie technique capable d'exploiter les vulnérabilités affectives des personnes[37]. La seconde question, connexe à la première, est celle de savoir dans quelle mesure la capacité d'illusion propre à ces technologies et l'asymétrie qui existera entre ces robots et les personnes dont ils analyseront les émotions les rendent moralement acceptables ? Sherry Turckle, professeure au MIT, souligne ainsi que les êtres humains ont une grande propension à attribuer aux robots une subjectivité et une sensibilité[38]. Or, la tentation est forte pour des sociétés vieillissantes de confier de plus en plus le soin des personnes âgées à ce type de robots. En France, Serge Tisseron développe une réflexion critique sur ces technologies[39] . Quelles que soient les réponses apportées à ces questions, il semble essentiel qu'elles n'occultent nullement la dimension politique et de choix de société que recèle le fait de recourir aux robots plutôt que d'investir dans d'autres types de ressources (temps, ressources en personnel, etc.) pour l'accompagnement des membres vulnérables de nos sociétés.

35 La question de l'hybridation entre l'homme et des artefacts n'est pas nouvelle : les algorithmes participent au modelage de notre identité de la même façon que – Socrate le remarquait déjà dans le *Phèdre* de Platon – l'écriture affecte notre capacité de mémorisation et constitue un artefact muet, incapable de la moindre explication. Que l'idée d'une « identité humaine » strictement distincte des objets soit remise en cause n'implique ainsi pas nécessairement une nouveauté radicale.
36 Gérard Berry, « Non, l'intelligence artificielle ne menace pas l'humanité ! », interview donnée au Point, 18 mai 2015.
37 Une problématique très similaire à celle soulevée par les logiciels de communication politique censés adapter le message du candidat aux attentes de chaque individu ciblé et profilé.
38 Sherry Turckle, *Seuls ensemble,* Paris, L'Echappée, 2015 [2012].
39 Serge Tisseron, *Le Jour où mon robot m'aimera. Vers l'empathie artificielle,* Paris, 2015.

# Quelles réponses ?

# Quelles réponses ?

## De la réflexion éthique à la régulation des algorithmes

### Faut-il réguler les algorithmes ?

La question se trouve depuis quelques mois fréquemment évoquée aussi bien dans la presse généraliste que parmi les experts du numérique et des politiques publiques.
Elle ne constitue que le prolongement de la question de la régulation du numérique lui-même. On le sait, l'univers numérique s'est constitué en partie en opposition à l'idée même de normes, du moins de normes juridiques. De la contre-culture américaine des années 1960 à la mise en avant par les entreprises numériques de la nécessité de ne pas entraver l'innovation par un système de normes inadaptées à un univers fondamentalement nouveau, cette méfiance à l'égard de la régulation trace comme un fil rouge. Ce courant de pensée a trouvé une de ses manifestations les plus claires dans la fameuse Déclaration d'indépendance du cyberespace de John Perry Barlow en 1996. Il se heurte depuis de nombreuses années aux efforts déployés par les acteurs étatiques pour soumettre l'univers numérique au droit commun, parfois de manière mécanique, parfois en mettant en œuvre de véritables innovations juridiques.

**De nombreux acteurs expriment aujourd'hui l'idée qu'il ne faudrait pas réguler les algorithmes et l'intelligence artificielle.** Ces derniers soulignent en effet qu'il serait trop tôt pour imposer des règles qui s'avéreraient nécessairement inadaptées et vouées à être rendues rapidement caduques par des évolutions techniques progressant désormais à un rythme incommensurable à celui de l'invention juridique.

Une telle position néglige à vrai dire une réalité juridique aussi massive que parfois inaperçue : **les algorithmes et leurs usages se trouvent d'ores et déjà encadrés, directement ou indirectement, par de nombreuses règles juridiques.** Il est vrai que ces règles, comme on le verra, se trouvent en fait dispersées dans divers lois et codes, à la mesure de la transversalité du numérique.

Par ailleurs, des sondages effectués à l'occasion du débat public initié par la CNIL ont mis en évidence une attente de règles et de limites en matière d'algorithmes et d'intelligence artificielle. Ces règles et ces limites peuvent être conçues autrement que comme des normes contraignantes, par exemple sous la forme de « chartes » adoptées par une entreprise, par une profession, par une branche. C'est ce que montre par exemple le sondage réalisé par la CFE-CGC auprès de 1263 de ses adhérents[40].

La création par le Parlement d'une mission de réflexion confiée à la CNIL sur les enjeux éthiques et de société soulevés par l'évolution des technologies numériques s'inscrit dans ce contexte. Elle traduit évidemment un souci de réflexion sur les limites, sur les normes – quelle que soit la nature de ces dernières – à imposer à des nouveautés techniques. Elle traduit tout autant une volonté de la part de la puissance publique de ne pas céder à la tentation de réguler trop vite et de manière inadaptée. À cet égard, considérer que l'émergence et la diffusion de technologies nouvelles implique une réflexion sur ses limites ne signifie nullement que la loi soit systématiquement la forme adaptée à l'imposition de ces limites. C'est en tout cas ce qu'a considéré la CNIL en souhaitant ouvrir la réflexion de la façon la plus large possible, non seulement aux acteurs publics mais aussi aux praticiens, professionnels et grand public.

Formuler des recommandations impliquait donc d'abord d'explorer les grands développements des innovations considérées et les enjeux éthiques et de société soulevés par ceux-ci. Les pages précédentes y ont été consacrées. Les pages suivantes viseront à faire le point sur les grands principes susceptibles de répondre à ces enjeux ainsi que sur les recommandations concrètes envisageables aujourd'hui.

**40** À la question « La définition d'une charte éthique autour de l'usage des algorithmes dans le recrutement et la gestion RH vous semble-t-elle une priorité ? », 92% ont répondu positivement.

# Ce que la loi dit déjà sur les algorithmes et l'intelligence artificielle

Tous les défis identifiés dans le présent rapport ne sont pas nouveaux.

La Commission Tricot, dont le rapport a constitué la base de la loi de 1978 sur la protection des données à caractère personnel, en avait déjà identifié certains à l'issue d'une réflexion qui, au-delà du traitement des données, portait sur les défis soulevés par l'informatisation de l'État et de la société française. Le risque de discrimination ou d'exclusion des personnes mais également le risque d'une confiance excessive accordée à l'ordinateur étaient d'emblée clairement identifiés, à côté des enjeux directement liés à la capacité de collecter et de stocker de grandes quantités de données. Les débats portant sur la nécessité ou non de « réguler les algorithmes » ignorent en fait purement et simplement le fait que les algorithmes sont encadrés par la loi (loi Informatique et Libertés, notamment, mais pas seulement) depuis une quarantaine d'années.

> Les débats portant sur la nécessité ou non de « réguler les algorithmes » ignorent le fait que les algorithmes sont encadrés par la loi depuis une quarantaine d'années

Aboutissement du travail de la Commission Tricot, la loi Informatique et Libertés de 1978 contient en effet un certain nombre de dispositions que l'on peut, de façon schématique, rattacher à trois principes, eux-mêmes abrités sous un principe général contenu dans l'article 1 : « l'informatique doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques ».

Ces trois principes se trouvent relayés dans le Règlement européen sur la protection des données personnelles (RGPD) entrant en vigueur en mai 2018. Ils sont les suivants :

**Premièrement, la loi encadre l'utilisation des données personnelles nécessaires au fonctionnement des algorithmes,** au-delà même du traitement algorithmique à proprement parler. Autrement dit, elle encadre les conditions de collecte et de conservation des données[41], ainsi que l'exercice de leurs droits par les personnes (droit à l'information, droit d'opposition, droit d'accès, droit de rectification) afin de protéger leur vie privée et leurs libertés.

**Deuxièmement, la loi Informatique et Libertés interdit qu'une machine puisse prendre seule (sans intervention humaine) des décisions** emportant des conséquences cruciales pour les personnes (décision judiciaire, décision d'octroi de crédit, par exemple)[42].

**Troisièmement, la loi prévoit le droit pour les personnes d'obtenir, auprès de celui qui en responsable, des informations sur la logique de fonctionnement de l'algorithme[43].**

Au-delà de la loi Informatique et Libertés, d'autres dispositions légales plus anciennes constituent de fait un cadre et une série de limites à l'utilisation des algorithmes dans certains secteurs, dans la mesure même où ils régulent ces secteurs[44]. La question de la collusion algorithmique qui se pose aujourd'hui aux régulateurs de la concurrence, par exemple, ne se pose pas dans un vide juridique : elle a plutôt trait à l'effectivité de la règle de droit et à la nécessité d'inventer de nouveaux moyens de prouver l'existence d'ententes illégales[45].

Les dispositions juridiques interdisant différentes formes de discrimination, élaborées dans le sillage de l'article 7 de la Déclaration universelle des droits de l'homme, s'appliquent naturellement aux algorithmes[46].

41 Principes de finalité, de proportionnalité, de sécurité, de limitation de la durée de conservation des données.
42 Article 10 de la loi de 1978. Article 22 du RGPD.
43 Article 39 de la loi de 1978. L'article 15.1 (h) du Règlement européen sur la protection des données personnelles (RGPD) prévoit que la personne peut obtenir du responsable de traitement des informations concernant "the existence of automated decision making including profiling referred to in Article 20(1) and (3) and at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject". Les limites juridiques posées par le RGPD portent notamment sur « le profilage » (pas de décision basée uniquement sur un traitement sauf exceptions).
44 On pourrait envisager, en forçant un peu le raisonnement, l'application du code de la santé publique (qui réprime l'exercice illégal de la médecine par toute personne non titulaire d'un diplôme) à des dispositifs d'intelligence artificielle dans le domaine médical. On pourrait imaginer qu'une telle disposition aussi puisse fonder l'interdiction de l'établissement d'un diagnostic par un algorithme seul. L'origine de cette législation, au début du XIXe siècle, renvoie à la préoccupation des autorités de lutter contre le « charlatanisme ». Les critiques des promesses excessives portées par certaines entreprises y verront sans doute un écho plaisant à la situation actuelle.
45 http://internetactu.blog.lemonde.fr/2017/02/11/comment-prouver-les-pratiques-anticoncurrentielles-a-lheure-de-leur-optimisation-algorithmique/
46 « Tous sont égaux devant la loi et ont droit sans distinction à une égale protection de la loi. Tous ont droit à une protection égale contre toute discrimination qui violerait la présente Déclaration et contre toute provocation à une telle discrimination ».

# Les limites de l'encadrement juridique actuel

Un certain nombre des enjeux soulevés par les algorithmes constituent cependant à ce jour un angle mort du droit et des différentes dispositions juridiques évoquées précédemment.

## Focalisation sur les algorithmes traitant des données personnelles et absence de prise en compte des effets collectifs des algorithmes

En premier lieu **ces dispositions ne concernent les algorithmes que dans la mesure où ils utilisent pour fonctionner des données à caractère personnel** et où leurs résultats s'appliquent directement à des personnes. C'est notamment le cas de la loi Informatique et libertés, la seule parmi celles évoquées qui vise directement les algorithmes (mentionnés comme « traitement automatisés de données à caractère personnel »). Or, bien des algorithmes n'utilisent pas de données à caractère personnel. C'est par exemple le cas des algorithmes boursiers. Les impacts de ces algorithmes traitant des données non personnelles sont tout aussi susceptibles que les autres de soulever des questions. Si les algorithmes boursiers relèvent d'un secteur par ailleurs fortement encadré, d'autres exemples peuvent permettre de comprendre les impacts que peuvent avoir des algorithmes ne traitant pas des données à caractère personnel. Celui, déjà évoqué au début de ce rapport (Voir « Une question d'échelle : la délégation massive de décisions non critiques »), de l'algorithme imaginé par Cathy O'Neil pour composer les repas de ses enfants lui permet de mettre en lumière les enjeux spécifiques liés à l'échelle de l'impact des algorithmes exécutés par des systèmes informatiques. On pourrait imaginer également un algorithme établissant les menus des cantines scolaires selon certains critères (optimisation du coût des denrées, qualité nutritionnelle, etc.) et qui pourrait être utilisé à l'échelle d'un pays. Un tel algorithme, sans traiter de données personnelles, serait susceptible d'avoir des impacts sociaux et économiques du fait même de son échelle de déploiement. Or, la loi n'a jusqu'ici pas pris en compte cette dimension nouvelle.

En second lieu, **les dispositions légales évoquées précédemment concernent les effets des algorithmes sur les personnes, dans une perspective individualiste. En revanche, elles ne visent pas directement leurs effets sur des collectifs.** Nous pensons ici par exemple aux impacts des algorithmes utilisés à des fins de marketing électoral sur le fonctionnement démocratique même (Voir : « Atomisation de la communauté politique »). Si l'on peut considérer que la loi Informatique et libertés constitue de fait un facteur limitant de tels impacts[47], ce n'est cependant que de manière indirecte, sans que ce soit son objectif premier.

## Les limites de l'effectivité du droit

**Un autre type de limites de l'encadrement des algorithmes et de l'IA identifiable dans les dispositions juridiques évoquées a trait à l'effectivité même de ces dernières et des principes qu'elles ont vocation à mettre en œuvre.** Dans un univers numérique caractérisé par une fluidité et une omniprésence des capteurs rendant difficile l'exercice des droits ainsi que par une forte asymétrie entre ceux qui contrôlent algorithmes et données et les personnes, ces dernières rencontrent des difficultés à exercer leurs droits (par exemple, le droit d'obtenir une intervention humaine dans le cadre d'une décision prise sur le fondement d'un traitement algorithmique, ou encore le droit d'obtenir une information sur la logique sous-tendant le fonctionnement de l'algorithme).

La prise en compte de cette réalité s'est traduite par une série de réflexions récentes, dont certaines se sont traduites dans de nouvelles dispositions légales. Le Règlement européen sur la protection des données à caractère personnel (entrée en application en mai 2018) apporte plusieurs réponses à cette question de l'effectivité du droit dans l'univers numérique, y compris en ce qui concerne les algorithmes[48]. Par ailleurs, la loi pour une République numérique (adoptée en octobre 2016) s'est inscrite dans cette même perspective de renforcement de l'effectivité de principes préexistants.
D'une part, elle a renforcé l'obligation faite à ceux qui déploient des algorithmes d'en informer les personnes concernées. D'autre part, elle prévoit que les codes sources des algorithmes utilisés par l'administration sont des documents communicables, approfondissant ainsi (à l'exception notable du secteur privé) le droit d'obtenir des informations sur la logique mise en œuvre par un algorithme présent dans la loi de 1978.

**47** La loi Informatique et libertés limite conditionne notamment au consentement des personnes l'enrichissement de profils individuels par des données collectées sur les réseaux sociaux.
**48** L'article 14.1a du Règlement européen, par exemple, renforce le droit à l'information, en prévoyant une information claire et intelligible fournie spontanément par le responsable du traitement algorithmique.

# Faut-il interdire les algorithmes et l'intelligence artificielle dans certains secteurs ?

La question de savoir s'il faut interdire les algorithmes et l'intelligence artificielle dans certains secteurs ou pour certains usages ne saurait être éludée d'une réflexion sur les enjeux éthiques soulevés par ces technologies. Rand Hindi évoquait ainsi lors de l'événement de lancement du débat organisé à la CNIL le 23 janvier 2017 la question de savoir s'il faudrait refuser d'automatiser certains métiers pour des raisons éthiques.

**Le caractère particulièrement sensible d'un certain nombre de secteurs et des décisions qui y sont prises les désigne assez logiquement comme étant ceux où la question de telles interdictions pourrait se poser.** Ainsi, le secteur militaire a récemment fait l'objet d'une pétition internationale demandant que soient bannies les armes autonomes. La médecine ou la justice constituent d'autres domaines où la question pourrait être posée. Certes, comme cela a été rappelé, la législation prévoit d'ores et déjà que le diagnostic du médecin ou la décision du juge ne puissent faire l'objet d'une automatisation. Devant le caractère toujours incertain de la frontière entre délégation et aide à la décision, la question d'un rappel solennel de ces principes pourrait être posée.

Certains secteurs à la sensibilité moins immédiatement évidente font également l'objet de demandes d'interdiction. Ainsi, Serge Tisseron a récemment pris position contre le ciblage personnalisé dans le domaine publicitaire et culturel, accusé de « condamner chaque spectateur à tourner en rond dans ce qu'il connaît de ses goûts et ce qu'il ignore de ses a priori » et de contribuer à « réduire les données dont la majorité des humains disposent pour se faire une opinion sur le monde[49] ».

Enfin, l'interdiction appliquée à tel ou tel usage des algorithmes pourrait porter sur les données utilisées, à l'image du moratoire mis en place par les assureurs français dès 1994 sur le recours aux données génétiques, relayé en 2002 par la loi Kouchner. Dans ce même secteur, une limitation du recours aux données ne serait-il pas une solution possible (légale ou mise en place par les acteurs eux-mêmes) pour maintenir le « voile d'ignorance indispensable » à la pérennité de la mutualisation du risque ?

## LE REGARD DU CITOYEN

Les participants à la concertation citoyenne organisée par la CNIL à Montpellier le 14 octobre 2017 (voir « L'organisation du débat public sur les enjeux éthiques des algorithmes et de l'intelligence artificielle ») ont identifié un certain nombre d'enjeux éthiques soulevés par les algorithmes et l'intelligence artificielle. Si leur positionnement révèle des inquiétudes et une conscience des risques, leur attitude générale ne traduit guère d'hostilité de principe à ce que des algorithmes et des outils d'intelligence artificielle se déploient dans notre quotidien, *sous réserve que des réponses soient apportées*.

Parmi les avantages mentionnés dans les différents ateliers de la journée de concertation, figurent la personnalisation du diagnostic médical, la fluidification de processus de recrutement qui deviendraient plus neutres, la simplification de la répartition des étudiants par rapport à l'offre de formation (APB) ou encore l'utilité des filtres sur les plateformes en ligne pour gérer « la multitude d'informations ».

Beaucoup voient positivement les capacités nouvelles d'analyse des données : 63% considèrent ainsi utile de « partager les données pour le bien commun ».

La montée en compétence des participants au cours de la journée de concertation se traduit par un certain accroissement de la conscience des risques : 32% des participants les considéraient comme « plutôt source d'erreur » à l'issue de la journée alors qu'ils n'étaient que 23% ex-ante. Une évolution certes modérée à l'issue d'une journée consacrée aux enjeux éthiques mais qui s'accompagne aussi d'une forme de scepticisme quant à la possibilité d'un encadrement effectif des algorithmes : « est-ce que la loi sera suffisante pour tout contrôler ? Ne sera-t-on pas toujours dans la correction après dérive ? ».

49 http://www.huffingtonpost.fr/serge-tisseron/les-publicites-ciblees-cest-la-betise-assuree-interdisons-les_a_23220999/

# Deux principes fondateurs pour le développement des algorithmes et de l'intelligence artificielle : loyauté et vigilance

La réflexion sur les enjeux éthiques soulevés par les algorithmes et l'IA a pour horizon deux dimensions distinctes mais articulées : les principes et les moyens concrets de rendre ceux-ci effectifs.

Le législateur avait inscrit à l'article 1 de la loi Informatique et Libertés que « l'informatique doit être au service de chaque citoyen ». Il s'agit aujourd'hui d'établir les principes permettant d'atteindre cet objectif général et de garantir que l'intelligence artificielle soit au service de l'homme, qu'elle l'augmente plutôt que de prétendre le supplanter.

Les principes inscrits dans la loi Informatique et Libertés et que l'on a rappelés précédemment correspondent-ils toujours aux enjeux qui ont été identifiés et à cet objectif général ? Faut-il en promouvoir de nouveaux ? Outre le constat que ces principes ne couvrent pas la totalité du champ des algorithmes et de l'IA, la circulation dans le débat public d'une série de notions représentant autant d'exigences à l'égard des algorithmes (loyauté, redevabilité, intelligibilité, explicabilité, transparence, etc.) signale à l'évidence le sentiment d'une inadéquation, voire d'inquiétudes.

Au terme du débat public, sont ici présentés une série de principes. Parmi ces derniers, deux en particulier, celui de loyauté et celui de vigilance, apparaissent comme tout particulièrement fondateurs.

## Le principe de loyauté

### Un principe formulé par le Conseil d'État

Dans son étude annuelle de 2014 sur le numérique et les droits fondamentaux, le Conseil d'État a ainsi formulé trois recommandations invitant à « repenser les principes fondant la protection des droits fondamentaux ». Parmi celles-ci, la première portait sur un principe d' « autodétermination informationnelle » garantissant la maîtrise de l'individu sur la communication et l'utilisation de ses données personnelles et depuis introduit dans la loi pour une République numérique. La troisième portait, elle, sur le principe de « loyauté », appliqué non pas à tous les algorithmes mais, de manière plus restreinte, aux « plateformes[50] ». Selon le Conseil d'État, « la loyauté consiste à assurer de bonne foi le service de classement ou de référencement, sans chercher à l'altérer ou à le détourner à des fins étrangères à l'intérêt des utilisateurs[51] ».

Parmi les obligations des plateformes envers leurs utilisateurs découlant du principe de loyauté tel que défini par le Conseil d'État figurent notamment, d'une part, la pertinence des critères de classement et de référencement mis en œuvre par la plateforme au regard de l'objectif de meilleur service rendu à l'utilisateur et, d'autre part, l'information sur les critères de classement et de référencement mis en œuvre. La première obligation pose donc une limite à la liberté d'établissement des critères de l'algorithme par la plateforme. La deuxième obligation fait de l'information sur la logique de fonctionnement de l'algorithme une obligation incombant à la plateforme (ce n'est pas seulement un droit que l'utilisateur peut choisir ou non de mobiliser).

Avec la loyauté ainsi définie, on accorde par ailleurs moins un droit aux utilisateurs qu'on impose une obligation à l'égard des responsables de traitement.

D'une certaine façon, le principe de loyauté se trouve sous une forme embryonnaire dans la loi Informatique et Libertés de 1978. En effet, le droit à l'information qui s'y trouve affirmé apparaît comme une exigence première de loyauté à l'égard de la personne concernée quant au fait même qu'un algorithme traite ses données. À cela s'ajoute le droit pour toute personne d'interroger le responsable du fonctionnement de l'algorithme pour obtenir des informations quant à la logique suivie par celui-ci ainsi que l'obligation de recueillir le consentement de la personne dont les données sont traitées. L'affirmation même de ces droits dans la loi de 1978 suppose que ces informations soient fournies de manière « loyale » et que le comportement de l'algorithme y corresponde effectivement.

L'intérêt du principe de loyauté tel qu'il est envisagé par le Conseil d'État réside dans la notion d' « intérêt des utilisateurs ». En effet, il ne s'agit pas simplement que l'algorithme dise ce qu'il fait et fasse ce qu'il dise : le principe de loyauté limite aussi la liberté que le responsable de l'algorithme a de déterminer les critères de fonctionnement de ce dernier. D'autre part, alors que dans la loi Informatique et Libertés,

50 Il s'agissait de « soumettre [les plateformes] à une obligation de loyauté envers leurs utilisateurs (les non professionnels dans le cadre du droit de la consommation et les professionnels dans le cadre du droit de la concurrence) ». Les plateformes apparaissent comme des acteurs classant un contenu qu'il n'a pas lui-même mis en ligne.
51 *Le Numérique et les droits fondamentaux,* 2014, p.273 et 278-281

l'information est un droit qui peut éventuellement être mobilisé par l'individu auprès du responsable de l'algorithme, avec le principe de loyauté, cette information doit d'emblée être diffusée à destination de la communauté des utilisateurs[52]. Il n'est pas question ici de droit des utilisateurs mais d'obligation des plateformes algorithmiques. Dans cette mesure, la loyauté semble à même de constituer une réponse au problème de l'asymétrie entre les responsables des algorithmes et les utilisateurs.

La notion de loyauté a notamment fait l'objet de réflexions complémentaires menées par le CNNUM. Celui-ci a en effet initié dans son rapport Ambition numérique (2015) une proposition tendant à créer une « agence de notation de la loyauté des algorithmes » appuyée sur un réseau ouvert de contributeurs, et ce dans un double objectif : rendre accessible via un point d'entrée unique toute une série d'informations déjà rassemblées par les différents acteurs ainsi que les outils existants et ouvrir un espace de signalement de pratiques problématiques ou de dysfonctionnements. Cette initiative pourrait, sous une forme ou sous une autre, participer à une meilleure connaissance citoyenne des enjeux, à une meilleure symétrie entre utilisateurs et plateformes algorithmiques, à une meilleure circulation des bonnes pratiques pour les entreprises ainsi qu'à un repérage facilité des pratiques litigieuses par le régulateur.

**Un principe à élargir pour prendre en compte les effets collectifs des algorithmes**

Toutefois, par rapport à la définition fournie par le Conseil d'État, **on peut estimer souhaitable d'élargir le principe, au-delà des seules plateformes, à tous les algorithmes**[53]. Par exemple, un algorithme d'aide à la décision en matière médicale, ne devrait-il pas faire l'objet d'une interdiction de recourir, ou du moins d'accorder une place excessive, à un critère lié à l'optimisation de l'occupation des lits d'un hôpital ?

Dès lors, le principe de loyauté des algorithmes aurait aussi l'intérêt de concerner des algorithmes ou des enjeux que ne touchent pas la législation sur la protection des données personnelles. Il concernerait en effet aussi les algorithmes ne procédant pas à un profilage de leurs utilisateurs à des fins de personnalisation de leurs résultats (par exemple, il voudrait pour un moteur de recherche qui ne fournirait pas des résultats profilés).

On pourrait enfin considérer l'opportunité de **reprendre la proposition du Conseil d'État en élargissant, ou du moins en précisant la notion d'« intérêt des utilisateurs », de façon à prendre en compte non seulement la dimension commerciale et économique de cet intérêt, mais également sa dimension collective.** Il s'agirait de considérer que les critères de l'algorithme doivent aussi ne pas entrer trop frontalement en opposition avec certains grands intérêts collectifs, liés notamment au troisième enjeu éthique évoqué précédemment. Ces intérêts collectifs peuvent être entendus de deux façons. D'une part, il peut s'agir de l'intérêt de catégories, de segments constitués par la logique même du big data et de l'analyse algorithmique (des groupes ad hoc, constitués par le croisement de certains traits), qui sont susceptibles de faire l'objet de formes de discriminations. Ces catégories font l'objet des réflexions actuelles portant sur la notion de « group privacy[54] ». D'autre part, cet intérêt collectif peut-être pensé comme celui d'une société tout entière. Par exemple, l'exposition à la diversité culturelle ou d'opinions pourrait être considérée comme liée à « l'intérêt des utilisateurs », entendus certes comme consommateurs mais aussi comme citoyens et parties prenantes d'une collectivité (il conviendrait d'ailleurs d'évoquer directement « l'intérêt des utilisateurs et des citoyens »).

> Les critères de l'algorithme doivent ne pas entrer trop frontalement en opposition avec certains grands intérêts collectifs

Le principe de loyauté des algorithmes, s'il constitue à l'évidence une réponse à d'importants enjeux, se heurte avec la montée en puissance des algorithmes de « machine learning » à une sérieuse difficulté. Ces algorithmes, on l'a vu, peuvent se comporter de façon problématique pour les droits des personnes, y compris à l'insu de leurs concepteurs (biais et discriminations cachés liés aux corrélations effectuées par le système). La notion de loyauté des concepteurs d'algorithmes (ce que l'on entend habituellement de fait par le vocable « loyauté des algorithmes ») perd une part de sa portée dès lors que l'algorithme se comporte d'une façon qui reste opaque à ces mêmes concepteurs. Il faudrait pouvoir parler, au sens propre, de loyauté des algorithmes (mais cela a-t-il un sens ?) ou bien s'assurer que l'algorithme ne se comportera pas d'une façon non souhaitable, sans que l'on soit bien en mesure de préciser a priori ce que l'on entend par ce « non souhaitable ». Autrement dit, **un algorithme loyal ne devrait pas avoir pour effet de susciter, de reproduire ou de renforcer quelque discrimination que ce soit, fût-ce à l'insu de ses**

52 « Sans méconnaître le secret industriel, les plateformes devraient expliquer à leurs utilisateurs la logique générale de leurs algorithmes et, le cas échéant, la manière dont les utilisateurs peuvent les paramétrer. »

53 Précisons – pour couper court à toute inutile querelle sémantique – que l'emploi de l'expression de « loyauté des algorithmes » ne revient pas à anthropomorphiser un fait technique (l'algorithme) mais est un raccourci pratique pour désigner la loyauté de ceux qui conçoivent et déploient l'algorithme.

54 Brent Mittelstadt, From individual to group privacy in Big Data analytics, B. Philos. Technol. (2017) 30: 475. https://doi.org/10.1007/s13347-017-0253-7

**concepteurs.** Cette dernière piste est donc plus large que les premières réflexions évoquées plus haut sur la notion de loyauté, développées avant tout en référence à des préoccupations d'ordre commerciales, concurrentielles, dans la perspective du développement de pratiques résolument déloyales destinées à obtenir un avantage en manipulant l'algorithme.

## Le principe de vigilance

Si le principe de loyauté apparaît comme un principe substantiel fondateur, le principe de vigilance constitue quant à lui un principe plus méthodologique qui doit orienter la façon dont nos sociétés modèlent les systèmes algorithmiques.

L'un des défis identifiés consiste dans **le caractère mouvant et évolutif des algorithmes à l'heure du *machine learning***. Cette caractéristique est renforcée par **l'échelle inédite de l'impact potentiel des algorithmes** exécutés par des programmes informatiques et donc de l'application d'un même modèle. Ceci accroît l'imprévisibilité, le caractère évolutif et potentiellement surprenant des algorithmes et de leurs effets. **Comment donc appréhender et encadrer un objet instable**, susceptible de générer des effets nouveaux au fur et à mesure de son déploiement et de son apprentissage, des effets imprévisibles au départ ?

La promotion d'un principe d'« obligation de vigilance » pourrait être une façon d'aborder ce défi en prévoyant la prise en compte par les concepteurs et ceux qui déploient l'intelligence artificielle de cette caractéristique inédite. Par ailleurs, ce principe d'obligation de vigilance viserait aussi à contrebalancer le phénomène de confiance excessive et de déresponsabilisation dont on a vu qu'il était favorisé par le caractère de boîte noire des algorithmes et de l'IA.

Enfin, **ce principe de vigilance doit avoir une signification collective.** Plus que d'algorithmes, sans doute faudrait-il parler de systèmes algorithmiques, de complexes et longues « chaînes algorithmiques » composées de multiples acteurs (du développeur à l'utilisateur final, en passant par la société ayant collecté les données utilisées pour l'apprentissage, le professionnel qui réalise cet apprentissage, par celui qui a acheté une solution de *machine learning* qu'il va ensuite déployer, etc.). Ce phénomène – semblable à celui qui peut se développer le long d'une chaîne de sous-traitance – favorise la dilution du sentiment de responsabilité, voire simplement de la conscience des impacts que peuvent générer ces outils. Par exemple, le data scientist, s'il occupe une position essentielle, en amont de la chaîne algorithmique, ne saurait détenir toutes les clés et ne possède pas nécessairement la vision d'ensemble de l'action collective dont il forme le premier maillon. Le Conseil National des Barreaux, dans le rapport remis à la CNIL, souligne pour sa part que « le sens de l'éthique du lieu de mise en œuvre du programme peut être très différent de celui du concepteur du programme ». En outre, l'informatique porte en elle-même le risque du développement d'une confiance exagérée dans une machine souvent perçue comme infaillible et exempte des biais charriés par l'action et le jugement humains. La commission Tricot, dans les années 1970, soulignait déjà ce risque. Plusieurs des intervenants du débat public l'ont mentionné cette année. Au total, donc, le développement des systèmes algorithmiques va de pair avec une érosion des vigilances individuelles. Or, il ne saurait être question de laisser se développer ce type d'indifférence face aux impacts possibles des algorithmes et de l'intelligence artificielle. Il est nécessaire d'organiser la vigilance collective, aussi bien à l'égard de phénomènes connus dont il s'agit d'éviter l'apparition qu'à l'égard de phénomènes ou d'impacts qui n'ont pas nécessairement pu être envisagés initialement mais dont l'échelle et le caractère évolutif des nouveaux algorithmes rendent la survenue toujours possible.

> Le développement des systèmes algorithmiques va de pair avec une érosion des vigilances individuelles

# Des principes d'ingénierie : intelligibilité, responsabilité, intervention humaine

## Intelligibilité, transparence, responsabilité

Face à l'opacité des systèmes algorithmiques, **la transparence** est une exigence très souvent affirmée, non sans lien d'ailleurs avec le principe de loyauté. Selon le Conseil national du numérique, « ce principe implique premièrement et d'une manière générale la transparence du comportement de la plateforme, condition pour s'assurer de la conformité entre la promesse affichée du service et les pratiques réelles. Dans les relations entre professionnels, il s'applique aux conditions économiques d'accès aux plateformes et aux conditions d'ouverture des services à des tiers[55] ». L'opacité en question concerne autant la collecte que le traitement des données par ces systèmes et donc le rôle que ceux-ci jouent dans un certain nombre de prises de décisions. Les algorithmes ne sont pourtant pas opaques seulement à l'égard de leurs utilisateurs finaux ou à ceux dont ils traitent les données. De plus en plus, avec l'affirmation du *machine learning,* les concepteurs mêmes de ces algorithmes probabilistes perdent la capacité à comprendre la logique des résultats produits. C'est donc à un double niveau que se pose la question de l'opacité. La transparence exigée face à cette situation appelle des réponses légales et de procédure mais elle soulève aussi un enjeu technique.

**L'idée de transparence des algorithmes est considérée par beaucoup comme excessivement simplificatrice et finalement insatisfaisante : une transparence assimilée à la publication pure et simple d'un code source laisserait l'immense majorité du public, non spécialisé, dans l'incompréhension de la logique à l'œuvre.** Par ailleurs, du moins en ce qui concerne le secteur privé, l'idée de transparence entre en tension avec le droit de la propriété intellectuelle, les algorithmes s'apparentant à un secret industriel dont la divulgation pourrait mettre en danger un modèle économique.

Enfin, des entreprises peuvent avancer de bonnes raisons de ne pas dévoiler le code source, ni les critères commandant le fonctionnement d'un algorithme. Ainsi Google cherche-t-il à éviter que les résultats fournis par l'algorithme de son moteur de recherche, PageRank, ne soient faussés par des acteurs qui seraient à même d'en exploiter la logique à leur profit.

De nombreux spécialistes proposent ainsi de préférer, à la transparence, l'exigence **d'intelligibilité** ou d'explicabilité des algorithmes. Plus que d'avoir accès directement au code source, l'essentiel serait d'être à même de comprendre la logique générale de fonctionnement de l'algorithme. Cette logique devrait pouvoir être comprise par tous et donc énoncée verbalement et non sous la forme de lignes de code. C'est ainsi la position de Daniel Le Métayer, de l'Institut national de recherche en informatique et en automatique (INRIA), pour qui l'intelligibilité passe à travers le questionnement sur la logique globale de l'algorithme ainsi que sur des résultats particuliers. C'est la position de Dominique Cardon : « 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. Par exemple, dans le marché du « search engine optimization », des acteurs cherchent à influer sur les résultats de 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 » (CNIL, événement de lancement du débat public, 23 janvier 2017).

L'idée d'intelligibilité (ou explicabilité), comme celle de transparence, s'articule de toute façon avec le principe de loyauté, dont on peut considérer qu'elle est finalement une condition de déploiement.

Enfin, l'introduction d'une obligation de redevabilité ou d'organisation de la responsabilité pourrait constituer une réponse au phénomène de dilution de la responsabilité qu'ont tendance à favoriser les algorithmes et l'intelligence artificielle. Il s'agirait de prévoir que le déploiement d'un système algorithmique doit nécessairement donner lieu à une attribution explicite des responsabilités impliquées par son fonctionnement.

55 https://cnnumerique.fr/wp-content/uploads/2015/11/CNNum_Fiche_Loyaute-des-plateformes.pdf

## Repenser l'obligation d'intervention humaine dans la prise de décision algorithmique ?

On a vu que la loi de 1978 avait posé un principe d'interdiction de toute prise de décision entraînant des effets juridiques à l'égard d'une personne sur le seul fondement d'un traitement automatisé de données à caractère personnel (autrement dit : sur le seul fondement du résultat fourni par un algorithme analysant des données personnelles). Ce principe est repris dans le Règlement européen sur la protection des données à caractère personnel. Néanmoins, l'un et l'autre de ces textes, immédiatement après avoir affirmé ce principe, le vident en grande partie de sa substance par l'adjonction d'exceptions très larges[56].

Il semble par ailleurs que le recours des juridictions à l'article 10 de la loi de 1978 (dont il est ici question) soit devenu moins fréquent et que l'interprétation dudit article soit devenue moins stricte au cours des quarante dernières années[57]. Une évolution de la Loi Informatique et Libertés intervenue en 2004 a d'ailleurs facilité de fait la prise de décision automatisée, dans le secteur bancaire (credit scoring) par exemple : si l'intervention humaine dans le processus est toujours requise, celle-ci prend la forme d'un droit pour la personne concernée de demander à ce que, en cas de décision défavorable, celle-ci soit réexaminée par une personne. Intervention humaine, donc, mais a posteriori et seulement sur demande.

Sans que le terme implique un jugement de valeur, **il semble que l'on puisse parler d'une forme de « dérive » ou d'évolution du seuil de tolérance de la société à l'égard de la prise de décisions automatisée depuis les années 1970.** L'évolution du droit et de la jurisprudence seraient le reflet de cette évolution. Ne faut-il pas dès lors revisiter le principe interdisant la prise de décision par une machine seule et impliquant donc la nécessaire intervention humaine ? Le revisiter pour accueillir les nouveaux usages de l'IA, sans toutefois y renoncer ?

Dans son étude annuelle de 2014, le Conseil d'État soulignait la nécessité d'assurer l'effectivité de l'intervention humaine. Il est possible de considérer qu'assurer l'effectivité de l'intervention humaine à l'échelle de chaque décision prise revient de fait à empêcher ou à limiter certaines applications des algorithmes et de l'IA. En effet, quand l'automatisation a pour fonction d'optimiser et d'accélérer un processus en remplaçant l'homme, une intervention humaine réellement effective pour chaque décision risque d'être dissuasive. On pourrait en fait poser ainsi la question : comment faire assurer par des machines des tâches auparavant accomplies par l'intelligence humaine (c'est la définition de l'IA) sans évacuer l'homme ? Une façon d'y répondre consiste à avancer que l'on pourrait envisager l'effectivité de l'intervention humaine autrement qu'à l'échelle de chaque décision individuelle. On pourrait, par exemple, assurer que des formes de délibération humaine et contradictoire encadrent et accompagnent l'utilisation des algorithmes en examinant et en interrogeant le paramétrage mais aussi tous les effets – directs et indirects – du système. Cette supervision pourrait ainsi porter, non pas sur chaque décision individuelle, mais de loin en loin sur des séries plus ou moins nombreuses de décisions.

La protection des libertés serait dès lors pensée moins en termes individuels que collectifs. On voit d'ailleurs ici comment une telle piste s'articulerait aussi avec l'idée d'une obligation de vigilance évoquée précédemment. Ce passage d'une interprétation individuelle à une interprétation collective de l'obligation d'assurer une forme d'intervention humaine dans la décision automatisée pourrait faire l'objet d'une modulation en fonction de la sensibilité des applications considérées et de la configuration de la balance avantages/risques (par exemple, dans la santé, faut-il considérer que la sensibilité des enjeux dépasse les gains et justifie donc un maintien de l'obligation de garantir une intervention humaine pour chaque décision ?).

> Comment faire assurer par des machines des tâches auparavant accomplies par l'intelligence humaine (c'est la définition de l'IA) sans évacuer l'homme ?

56 Sur ce point dans le RGPD, voir par exemple : Wachter, Sandra, Brent Mittelstadt, et Luciano Floridi. « Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation ». Social Science Research Network, décembre 2016

57 Voir par exemple la délibération de la CNIL sur le projet GAMIN, en 1981 : la Commission rejeta alors ce projet du ministère de la santé. Même les garanties pourtant données par le ministère pour assurer une intervention humaine effective dans la détection de mineurs à risques psycho-sociaux dont il était question furent repoussées. On peut pourtant se demander à la lecture du dossier si la position de la CNIL serait aujourd'hui la même, alors qu'il nous semble qu'une certaine accoutumance s'est opérée à l'idée de voir des algorithmes intervenir de plus en plus fortement dans des domaines de plus en plus importants. Par exemple, la décision d'éliminer des candidats sur le fondement d'un seul traitement automatisé ne paraît guère relever de la science-fiction ni même probablement de ce que beaucoup dans notre société sont prêts à accepter.

# Des principes aux recommandations pratiques

Comment donner une effectivité concrète aux principes abordés précédemment ? Les pages suivantes listent les principales recommandations qui ont émergé du débat public organisé par la CNIL de janvier à octobre 2017, complété par la consultation de rapports déjà émis par diverses institutions en France et à l'étranger (entre autres, l'OPECST, la CERNA, le CNNUM, le Conseil d'État, la CGE, la Maison Blanche, France IA, INRIA, AI Now).

Une idée générale qui émane de la plupart des réflexions est que les solutions impliquent nécessairement une palette d'actions diversifiées concernant différents acteurs (les concepteurs d'algorithmes, les professionnels, les entreprises, la puissance publique, la société civile, l'utilisateur final). Les systèmes algorithmiques et d'intelligence artificielle sont des objets socio-techniques complexes, modelés et manipulés par de longues et complexes chaînes d'acteurs. **C'est donc tout au long de la chaîne algorithmique** (du concepteur à l'utilisateur final, en passant par ceux qui entraînent les systèmes et par ceux qui les déploient) **qu'il faut agir, au moyen d'une combinaison d'approches techniques et organisationnelles.** Les algorithmes sont partout : ils sont donc l'affaire de tous.

La loi ne saurait être le seul levier, la solution passant nécessairement par une mobilisation de tous les acteurs. Un certain nombre des recommandations formulées ci-dessous ne précisent d'ailleurs pas si c'est la loi ou l'initiative spontanée des différents acteurs qui devrait être privilégiée pour les mettre en œuvre.

## LE REGARD DU CITOYEN

Les participants à la concertation citoyenne organisée par la CNIL à Montpellier le 14 octobre 2017 (voir « L'organisation du débat public sur les enjeux éthiques des algorithmes et de l'intelligence artificielle ») ont formulé des recommandations. Ces dernières recoupent en grande partie celles recueillies ailleurs dans le cadre du débat public.

- Le souhait que l'humain garde le contrôle sur le développement des algorithmes apparaît prioritaire (95% d'avis favorables), une délégation excessive des décisions aux algorithmes et à l'IA étant jugée préjudiciable. Le constat des participants rejoint l'idée d'un principe de vigilance précédemment évoqué : 97% souhaitent « garder la dimension humaine, garder une dose de subjectivité, ne pas de désinvestir totalement » et 91% considèrent que « l'utilisateur devrait être dans une posture d'apprenant à chaque usage d'un algorithme afin d'en cerner les limites et être exigeant vis-à-vis des développeurs à chaque fois que cela est nécessaire ». Dans le champ de la médecine par exemple, certains citoyens pensent qu'une partie des décisions devrait toujours être discutée en collège.

- L'adaptation de la formation des concepteurs d'algorithmes est une option ayant émergé dans plusieurs groupes de travail, et ayant fait l'objet d'un quasi-consensus : 97% des participants considèrent que « les développeurs doivent intégrer dans leurs pratiques une certaine éthique et résister aux demandes tentantes du marché qui peuvent affecter cette dimension ». 94% en appellent ainsi au développement de chartes éthiques et 56% souhaiteraient que des experts associés issus des sciences humaines et sociales permettent aux développeurs de mieux mesurer l'impact de leur travail sur la société. La formation concerne également les utilisateurs d'algorithmes : 82% des personnes présentes sont favorables à une obligation de formation continue des médecins utilisant des systèmes d'aide à la décision. Plus généralement, ce besoin de savoir et de comprendre se matérialise par une forte demande pour plus d'éducation au numérique tout au long de la vie. Notamment afin de lutter contre les inégalités face à ces objets, l'intégralité des citoyens revendiquent une « éducation populaire au numérique » et le « développement de programmes scolaires pour une « alphabétisation » au numérique tant sur l'objet que sur les enjeux ».

## LE REGARD DU CITOYEN (suite)

- La nécessité de disposer de droits renforcés en matière d'information, de transparence et d'explication quant à la logique de fonctionnement de l'algorithme a également été vigoureusement affirmée dans chacun des groupes de travail. C'est déjà la possibilité d'être informé dès qu'un algorithme est déployé qui semble être exigée par les participants : 88% des participants estiment en effet qu'un employeur qui utilise un algorithme devrait impérativement l'indiquer aux candidats. La mise à disposition des codes source est jugée souhaitable par 78% d'entre eux, bien qu'elle soit considérée comme insuffisante pour comprendre les résultats produits par un algorithme. Dans le cas d'APB par exemple, un accompagnement plus effectif pour comprendre les ressorts de son utilisation est demandé par 78% des citoyens présents. 85% voient d'ailleurs dans l'expérience des utilisateurs un matériau précieux pour « améliorer l'ergonomie de la procédure ». Lorsqu'un critère de l'algorithme repose sur des choix politiques (le tirage au sort, par exemple), il convient également de ne pas l'occulter mais bien au contraire de le rendre lisible (selon 94% des participants). Notons que si un désir de transparence se manifeste, il n'est pas unanime et s'accompagne d'une lucidité voire d'un fatalisme sur l'hypothèse qu'elle puisse suffire.

- Un effort étatique de régulation pour identifier les biais, « éviter les abus, établir des statistiques et imposer un agrément » pourrait constituer une solution selon une écrasante majorité des participants (97%). Beaucoup préconisent la création d'un organisme indépendant pour effectuer des tests scientifiques sur les algorithmes, « à l'image des médicaments avant la mise en vente sur le marché » (84%). Sur le long terme, s'assurer régulièrement que l'algorithme soit « toujours en phase avec les objectifs visés » constitue également une idée ayant émergé des débats (63% y sont favorables). L'intervention du législateur est également vivement souhaitée (94%) afin de mieux intégrer l'éthique dans les lois « à travers des chartes et des règles déontologiques, des formations, des concertations ».

- L'importance pour la société civile de s'organiser face à ces objets technologiques nouveaux a aussi été avancée par certains des participants à travers notamment le rôle du tissu associatif (associations de patients dans le champ de la santé), la protection nécessaire des lanceurs d'alerte, ou encore le soutien apporté à des réseaux alternatifs aux plateformes du web dont les algorithmes utilisés poseraient question.

- Enfin, les échanges ont démontré un fort attachement à la protection des données à caractère personnel et à la protection de la vie privée. La question de savoir à qui appartiennent nos données et quels sont les usages qui en sont faits a été jugée prioritaire dans certains groupes de travail, sur la santé notamment, ou encore sur l'emploi (inquiétude sur la possibilité que des algorithmes analysent des données qui seraient collectées en dehors de l'entreprise).

## RECOMMANDATION 1

### Former à l'éthique tous les maillons de la « chaîne algorithmique » : concepteurs, professionnels, citoyens

**Formation des citoyens**

Le citoyen est l'un des acteurs centraux des systèmes algorithmiques. D'une part car les algorithmes ont un impact croissant sur son existence. D'autre part, parce qu'il est particulièrement bien placé pour en identifier les éventuels dérives. Lui fournir les clés de compréhension lui permettant d'aborder de manière confiante, active et éclairée ces nouvelles technologies est une nécessité qui correspond par ailleurs à une demande de sa part, ainsi que l'a rappelé fortement la concertation citoyenne organisée à Montpellier par la CNIL le 14 octobre 2017.

L'impératif de constituer une « nouvelle littératie » numérique à intégrer dès l'école et jusqu'à l'université fait l'objet d'un large consensus. Des acteurs comme le CNNUM ont déjà développé des réflexions à cet égard[58]. Cette littératie numérique comprendrait évidemment une culture algorithmique dont les fondements peuvent d'ailleurs être posés très tôt, par des exercices qui n'impliquent pas nécessairement le recours à un matériel numérique.

58 https://cnnumerique.fr/education-2/

La diffusion de cette culture algorithmique très largement dans la population peut être également favorisée par l'encouragement aux initiatives de médiation numérique dans les territoires. Autrement dit, une forme d'éducation populaire numérique incluant une dimension d'appropriation aux données et aux algorithmes. Citons par exemple les initiatives de la FING (Info Lab), de la Péniche à Grenoble (Coop-Infolab), de Pop School à Lille.

**Formation des concepteurs d'algorithmes**

Les concepteurs d'algorithmes (développeurs, programmeurs, codeurs, data scientists, ingénieurs) forment le premier maillon de la chaîne algorithmique. Ils occupent à ce titre une position particulièrement sensible. La technicité de leurs métiers est par ailleurs susceptible de rendre leurs actions opaques (et donc difficilement contrôlables) aux autres acteurs. Il est capital qu'ils aient une conscience aussi claire que possible des implications éthiques et sociales de leurs actions, et du fait même que ces dernières peuvent recouvrir la dimension de choix de société qu'ils ne sauraient être légitimes à arbitrer seuls. Or, **l'organisation concrète du travail et de l'économie tend à segmenter les tâches et à favoriser la tendance que peuvent avoir les individus à ignorer les implications de leur activité au-delà de leur silo.** Par conséquent, il est nécessaire que leur formation même mette les concepteurs des algorithmes en capacité de saisir ces implications parfois très indirectes sur les personnes mais aussi sur la société, qu'elle les responsabilise en éveillant leur *vigilance*.

L'intégration, dans la formation des ingénieurs et data scientists, de l'approche des sciences humaines et sociales (sociologie, anthropologie, gestion, histoire des sciences et des techniques, sciences de l'information et de la communication, philosophie, éthique) sur ces questions peut à cet égard avoir des effets positifs.

Le développement de ces mêmes enseignements bénéficierait de l'intégration des approches techniques et des sciences humaines et sociales au sein de laboratoires interdisciplinaires.
Certaines initiatives vont déjà dans ce sens. Citons par exemple le cas de l'ENSC (École nationale supérieure de cognitique, à Bordeaux), grande école intégrant les sciences humaines et sociales au cursus de formation de ses ingénieurs ou encore le laboratoire Costech (Connaissance, organisation et systèmes techniques), à l'Université Technique de Compiègne (UTC).

Enfin, il est essentiel de favoriser la diversification culturelle, sociale et de genre des professions impliquées dans la conception des algorithmes afin de garantir que l'intelligence artificielle ne favorise pas des formes d'ethnocentrisme. La féminisation de ces métiers devrait notamment commencer par un effort d'ouverture des filières de formation aux femmes.

**Formation des professionnels utilisateurs d'algorithmes**

Pour considérer l'ensemble de la chaîne de déploiement des algorithmes, il est nécessaire d'envisager également la formation des professionnels appelés à utiliser ces systèmes dans le cadre de leur activité. Il s'agirait notamment de les armer contre le risque de déresponsabilisation, de perte d'autonomie que peut développer le recours à des outils fonctionnant parfois comme des boîtes noires présentées comme étant d'une efficacité imparable. Prévenir le développement d'une confiance excessive en sensibilisant aux dimensions éthiques d'une prise de décision qui ne doit pas exclure l'intervention humaine et en développant l'esprit critique s'avère crucial dans des secteurs particulièrement sensibles, comme peuvent l'être la médecine, le recrutement, la justice, et peut-être dès maintenant surtout le marketing, où les catégories antisémites récemment générées par les algorithmes apprenants de Facebook sont venus illustrer la réalité des risques. Cette formation devrait notamment inclure, dans une optique pluridisciplinaire, la prise en compte des enjeux spécifiques que posent ces outils à chaque secteur. Un médecin qui utilise un système d'aide au diagnostic recourant à l'intelligence artificielle, par exemple, devrait être rendu spécifiquement attentif au développement possible de biais et capable d'une réflexivité à la hauteur de l'outil qu'il maniera et des conséquences de ses erreurs.
On pourrait ainsi imaginer la création de sorte de « permis d'utiliser les algorithmes et l'IA » dans certains secteurs, acquis grâce à des modules de formations spécifiques que délivreraient universités et écoles spécialisées.

**Sensibilisation des acteurs publics à la nécessité d'un usage équilibré et « symétrique » des algorithmes**

De même, il serait souhaitable de sensibiliser les acteurs publics à la nécessité d'un déploiement équilibré et symétrique des algorithmes. Alors que ces derniers sont de plus en plus utilisés pour lutter la fraude et à des fins de contrôle, laisser se développer dans le public la perception erronée selon laquelle ils ne peuvent servir qu'au contrôle et à des finalités répressives (par ailleurs utiles aux individus mêmes) risquerait de générer une forme de défiance qui serait à terme néfaste à leur déploiement et à l'exploitation de leurs avantages. Il serait donc hautement souhaitable que les responsables administratifs et politiques soient convaincus de l'utilité d'exploiter les potentialités des algorithmes qui apparaissent immédiatement favorables aux personnes et permettent d'améliorer l'accès aux droits (détection du non-recours aux aides sociales)[59].

59 Dans une évaluation des politiques publiques en faveur de l'accès aux droits sociaux, les députés Gisèle Biémouret et M. Jean-Louis Costes proposaient en 2016 de « mettre les outils de lutte contre la fraude au service de la diminution du non recours aux droits sociaux ». Voir : Rapport d'information du comité d'évaluation et de contrôle des politiques publiques sur l'évaluation des politiques publiques en faveur de l'accès aux droits sociaux.

## RECOMMANDATION 2

### Rendre les systèmes algorithmiques compréhensibles en renforçant les droits existants et en organisant la médiation avec les utilisateurs

L'opacité, pour les personnes, des algorithmes qui les profilent et de la logique à laquelle ceux-ci obéissent, pour leur attribuer un crédit bancaire par exemple, n'est pas sans trouver de premiers éléments de réponse dans le droit existant. De même, on a vu que celui-ci contient depuis longtemps des dispositions ouvrant la voie à une première forme d'intelligibilité et de transparence[60].

En revanche, de nombreux diagnostics convergent pour souligner l'insuffisance de ces dispositions pour résorber de manière effective l'opacité des systèmes algorithmiques et assurer intelligibilité, transparence et loyauté. Instaurer pour les responsables des systèmes une obligation de communication (et non pas sur la seule demande formulée par les personnes concernées) claire et compréhensible des informations permettant de comprendre la logique de fonctionnement d'un algorithme serait une façon de répondre à cet enjeu. Elle a d'ailleurs été d'ores et déjà prévue dans la loi pour une République numérique pour les algorithmes déployés par les administrations publiques[61].

On peut également considérer qu'il serait souhaitable que cet impératif (qu'il soit fixé par la loi ou librement adopté par les acteurs) concerne aussi les algorithmes n'impliquant pas le traitement des données personnelles de leurs utilisateurs, dans la mesure où ceux-ci sont susceptibles d'avoir des impacts collectifs significatifs, même si ceux-ci ne portent pas directement sur des personnes (voir notamment « Les limites de l'encadrement juridique actuel des algorithmes » et « Le principe de loyauté »).

Une telle obligation, inscrite dans la loi, pourrait opportunément être prolongée par des initiatives privées enclenchant une dynamique vertueuse. Pour les acteurs du web ayant des sites sur lesquels les personnes disposent d'un compte auquel elles peuvent se connecter, l'information sur leur « profil », les données traitées et inférées et la logique de l'algorithme pourraient être accessibles dans cet espace. Les personnes pourraient par ce biais corriger et actualiser aisément leur profil et les données les concernant.

Cette évolution du droit pourrait être relayée par le développement de bonnes pratiques par les acteurs, à l'aide d'outils de droit souple.

Le problème de l'opacité des algorithmes tient aussi au fait que **les responsables des systèmes algorithmiques ne sont pas, dans l'immense majorité des cas, concrètement joignables ou accessibles pour fournir ces informations et explications.** Ceci implique également une irresponsabilité de systèmes auxquels les utilisateurs se trouvent dans l'impossibilité de demander des comptes. **Il est donc nécessaire d'organiser une forme de « joignabilité » des systèmes algorithmiques,** notamment en identifiant systématiquement au sein de chaque entreprise ou administration une équipe responsable du fonctionnement d'un algorithme dès lors que celui-ci traite les données de personnes physiques. Il est en outre nécessaire de communiquer délibérément et de façon claire l'identité et les coordonnées de cette personne ou de cette équipe de façon à ce qu'elle puisse être contactée aisément et qu'elle ait les moyens de répondre rapidement aux demandes reçues.

La joignabilité devrait être aussi accompagnée d'un effort résolu pour organiser la médiation et le dialogue entre les systèmes et la société, conformément aux idées développées par la Fondation Internet Nouvelle Génération (FING) dans le cadre de l'initiative « NosSystèmes ». La FING constate en effet que « joindre le responsable technique ne suffit pas ». Elle propose ainsi, par exemple, la mise en place d'équipes dédiées à la qualité du dialogue usager ainsi que d'un « pourcentage médiation ». Alors que les algorithmes permettent des économies d'échelle, prendre en compte le pourcentage du budget d'un projet consacré à l'effort de médiation (mise en place d'outils de visualisation, équipe de médiation, partenariat, contrôle de la bonne compréhension de l'information, etc.) pourrait permettre – via des procédures de certification – de valoriser et de conférer un avantage concurrentiel (en termes d'image aux yeux des consommateurs) aux systèmes vertueux.

## RECOMMANDATION 3

### Travailler le design des systèmes algorithmiques au service de la liberté humaine

Plus que l'algorithme seul, voire le programme exécutant l'algorithme, c'est à l'ensemble du système algorithmique qu'il s'agit de s'intéresser pour en comprendre et en contrôler les effets. De nombreuses réflexions récentes mettent en avant l'importance de prendre en compte le design des systèmes algorithmiques, c'est-à-dire l'interface entre la machine et son utilisateur.

**60** Notamment l'article 39 de la Loi Informatique et libertés, organisant le droit d'accès.
**61** L'article 14.1a du Règlement européen va dans ce sens en prévoyant une telle information.

Il convient ainsi d'agir sur le design pour contrer le caractère de « boîtes noires » que peuvent avoir les algorithmes dès lors qu'ils se présentent comme des systèmes opaques présentant des résultats sans mise en perspective de leurs propres limites ni présentation de la manière dont ils sont construits mais parés du prestige de la neutralité et de l'infaillibilité si facilement prêtées à la machine.

À l'inverse, il s'agit de promouvoir un design propre à renforcer l'autonomie et la réflexivité des personnes, à remédier aux situations d'asymétrie que peuvent établir les algorithmes à leur détriment, à leur permettre de prendre des décisions informées et de manière lucide.

Par exemple, la mise en place de systèmes de visualisation permettant de redonner plus de contrôle à l'utilisateur en lui donnant une meilleure information va dans ce sens. **Des outils de visualisation peuvent permettre à des individus de comprendre pourquoi des recommandations leur ont été proposées voire, encore mieux, de générer en retour des recommandations plus appropriées.** Les individus se trouvent par-là même placés dans une posture active. Il s'agit de donner à l'individu la main sur une partie au moins des critères qui déterminent la réponse fournie par l'algorithme, lui permettant éventuellement de tester différentes réponses en fonction de paramétrages différents. Un exemple d'outil de visualisation a été fourni au cours du débat public par la présentation du « Politoscope[62] ». Développé par l'Institut des Systèmes Complexes de Paris-Île de France, le politoscope permet au grand public de plonger dans des masses de données et de voir l'activité et la stratégie des communautés politiques sur les réseaux sociaux et notamment sur Twitter. Il aide à contrebalancer, en la dévoilant, la pratique de l'astroturfing, c'est-à-dire la manipulation à leur avantage par des groupes très organisés des réseaux sociaux pour imposer certains thèmes à l'ordre du jour de la scène politique nationale. Il participe ainsi à un rééquilibrage dans l'utilisation des algorithmes, dans le but de préserver un accès démocratique à l'information.

À travers le design, c'est toute la relation entre l'homme et la machine qui peut être modifiée, dans le sens d'une responsabilisation de l'homme et d'une augmentation de sa capacité à prendre des décisions éclairées, au lieu d'une confiscation au profit de la machine de sa capacité à faire des choix. C'est en somme au *principe de vigilance* évoqué plus haut qu'il s'agit ici de donner corps.

Le concept de « jouabilité » récemment proposé par la FING dans le cadre de son expédition « NosSystèmes » pourrait également constituer un principe régissant le design de systèmes algorithmiques vertueux, mis au service de l'individu et de sa capacité d'agir dans toute sa plénitude. Il s'agit de permettre aux utilisateurs de « jouer » avec les systèmes en en faisant varier les paramètres. Permettre par exemple aux utilisateurs d'APB de pouvoir le tester « à blanc » en voyant les résultats fournis en fonction de différents choix avant d'entrer leurs vœux définitifs. On pourrait ainsi imaginer également qu'un moteur de recherches sur internet permette à ses utilisateurs de lancer plusieurs recherches en faisant varier les critères. L'idée de jouabilité repose sur le fait que **toucher et manipuler est la clé d'une compréhension directe,** bien davantage sans doute que l'accès à un code source indéchiffrable pour la grande majorité d'entre nous.

> À travers le design, c'est toute la relation entre l'homme et la machine qui peut être modifiée, dans le sens d'une responsabilisation de l'homme et d'une augmentation de sa capacité à prendre des décisions éclairées

## RECOMMANDATION 4

### Constituer une plateforme nationale d'audit des algorithmes

Développer l'audit des algorithmes de manière à contrôler leur conformité à la loi et leur loyauté est une solution fréquemment évoquée pour assurer leur loyauté, leur responsabilité et, plus largement, leur conformité à la loi.

Développer l'audit des algorithmes signifie d'abord développer la capacité de la puissance publique à assurer ce dernier. L'audit des algorithmes n'est pas une réalité nouvelle. La Commission des sanctions de l'Autorité des marchés financiers s'est ainsi appuyée sur l'analyse de l'algorithme « Soap » pour rendre sa décision du 4 décembre 2015 à

62 https://politoscope.org/

l'égard des sociétés Euronext Paris SA et Virtu Financial Europe Ltd. De même, la CNIL dispose, pour son activité de contrôle, des compétences d'auditeurs des systèmes d'information. L'Autorité de la concurrence doit aussi appuyer de plus en plus son activité sur une capacité à auditer les algorithmes.

Il est donc essentiel que la puissance publique se donne autant que possible les moyens d'ouvrir le code source d'algorithmes déterministes. Or, ces moyens s'avèrent de plus en plus insuffisants face à un besoin croissant. La CNIL se trouve ainsi désormais sollicitée par d'autres régulateurs sectoriels dépourvus de toute capacité d'audit. **Un travail de recensement des ressources de l'État, des différents besoins ainsi qu'une mise en réseau des compétences et des moyens au moyen d'une plateforme nationale est donc aujourd'hui une nécessité.**

Une telle plateforme devrait aussi avoir pour fonction de relever le défi que soulève le développement du *machine learning.* Celui-ci conduit certains à souligner que l'examen des codes sources s'avère peu réaliste dès lors qu'il s'agit d'analyser des millions de lignes de code. Or, auditer ne signifie pas nécessairement ouvrir les codes sources. Cela peut aussi prendre la forme de contrôles ex post des résultats produits par les algorithmes, de tests aux moyens de profils fictifs, etc. Ces techniques d'audit reposant sur la rétro-ingénierie doivent faire l'objet d'un effort de recherche significatif (cf. recommandation suivante).

Opérationnellement, la mise en œuvre de ces audits pourrait être assurée par un corps public d'experts des algorithmes qui contrôleraient et testeraient les algorithmes (en vérifiant par exemple qu'ils n'opèrent pas de discrimination). Une autre solution pourrait consister, notamment face à l'ampleur du secteur à contrôler, à ce que la puissance publique homologue des entreprises d'audit privées sur la base d'un référentiel. Certaines initiatives privées ont d'ailleurs d'ores et déjà vu le jour. Par exemple, Cathy O'Neil, plusieurs fois citée dans ce rapport, a créé la société « Online Risk Consulting & Algorithmic Auditing », une entreprise dont l'objectif est d'aider les entreprises à identifier et à corriger les préjugés des algorithmes qu'elles utilisent.

Indépendamment même d'une obligation de recourir à la procédure d'audit, il est souhaitable que les entreprises et les administrations se tournent vers des solutions de type « label ». Ces labels pourraient alimenter une dynamique vertueuse. D'une part, ils garantiraient la non-discrimination et la loyauté des algorithmes. D'autre part, ils offriraient aussi une visibilité aux efforts en vue de la mise en place d'un design ainsi qu'en vue de la mise en place d'une information proactive et adaptée, conformément aux recommandations précédentes, donc, et au-delà même des strictes obligations légales.

## RECOMMANDATION 5

## Encourager la recherche de solutions techniques pour faire de la France le leader de l'IA éthique

**Favoriser l'explication sur le fonctionnement et la logique des algorithmes.**
Fournir aux régulateurs, aux entreprises et aux citoyens des outils robustes pour contrôler, maîtriser et surveiller les impacts des algorithmes et de l'intelligence artificielle, pour en comprendre la logique de fonctionnement devrait constituer un axe croissant des politiques de recherche.

Le développement de **techniques de rétro-ingénierie** pour « tester » le caractère non discriminatoire, la capacité à **pré-traiter les données pour réduire les risques de discrimination** en identifiant et résolvant les biais des jeux d'apprentissage[63], la **génération d'explications en langage naturel par les machines algorithmiques recourant à l'apprentissage automatique des résultats qu'elles produisent** mériteraient de faire l'objet d'un investissement significatif.

En France, le projet TransAlgo, conduit par INRIA, a d'ores et déjà pour objectif de catalyser la dynamique sur ces questions à travers l'élaboration d'une plateforme scientifique. Le projet Algodiv (recommandation algorithmique et diversité des informations du web) vise quant à lui à apporter des réponses aux questions posées par la notion d'enfermement : les algorithmes nuisent-ils à la diversité et à la sérendipité ? Ces projets ont en somme pour but de fournir une meilleure compréhension d'un certain nombre de problèmes évoqués dans le présent rapport.

Des initiatives visant à articuler interdisciplinarité, recherche de pointe et développement d'outils devraient être soutenues en France, à l'instar de celle du Professeur Katharina Anna Zweig en Allemagne qui a créé en 2017 le laboratoire « Algorithmic Accountability Lab ». Ce dernier, outre une activité d'analyse appuyée sur les sciences dures, les sciences techniques et les sciences humaines (conformément à l'idée que les systèmes algorithmiques ne peuvent être compris, prédits et contrôlés que dans le contexte de leur application), vise à développer un design transparent, éthique et responsable des systèmes automatisés d'aide à la décision. Il propose en outre des outils didactiques

63 La construction d'un jeu de données non-biaisées a fait cette année l'objet d'un projet mené par l'association Open Law, partenaire du débat public animé par la CNIL.

concernant les risques et promesses des ADM[64] pour, à la fois, le grand public et les preneurs de décision[65].

Un autre exemple est fourni par la constitution récente aux États-Unis de l'Institut de recherche *AI Now* (au sein de la New York University) dont l'objet est d'examiner les implications sociales de l'intelligence artificielle. L'implication dans cette structure du « Partnership on AI », initiative notamment d'Amazon, Apple, Google, IBM et Microsoft amène toutefois à souligner l'attention toute particulière qui devrait être attachée à la composition de telles institutions. Comme l'a récemment souligné l'ancienne universitaire Cathy O'Neil, l'importance de l'implication du monde de la recherche dans le travail d'élucidation des impacts sociaux de l'IA tient aussi à la valeur toute particulière de la liberté académique[66].

**Développer des infrastructures de recherche respectueuses des données personnelles**

Le développement d'une IA respectueuse des données constitue un enjeu croissant alors que les citoyens en Europe, mais aussi plus largement dans le monde entier, se montrent de plus en plus soucieux de la protection de leurs données personnelles et aux risques générés. Diverses pistes peuvent ici être évoquées dans la perspective de la construction d'un nouvel équilibre, fondé sur un renforcement symétrique des capacités d'accès des chercheurs à d'importants jeux de données et de la sécurité de ces mêmes données

Tout d'abord **le développement d'espaces sécurisés pour l'accès à des données** à des fins de recherche et d'entraînement des algorithmes d'intelligence artificielle. Par exemple, des travaux comme ceux conduits dans le cadre du projet OPAL, participent de cette dynamique. Ce projet vise à bâtir une infrastructure sur laquelle les données d'opérateurs téléphoniques sont stockées et peuvent être analysées en toute sécurité au moyen d'algorithmes certifiés mis à disposition des utilisateurs et enrichissables par la communauté. Avec de tels systèmes, les données ne sont pas directement accessibles à ceux qui les exploitent, garantissant la protection des personnes. La certification des algorithmes qui peuvent être utilisés pour analyser ces jeux de données a une fonction de filtrage éthique des données, ce qui permet notamment de faire face aux défis posés en termes de « *group privacy*[67] ».

**Des bases à la main d'acteurs publics** tels que le CASD (Centre d'Accès Sécurisé aux Données) utilisées pour la mise à disposition de bases de données de l'administration à des fins de recherche constituent également une piste à suivre.

**Lancer une grande cause nationale participative pour dynamiser la recherche en IA**

La capacité à disposer de très vastes quantités de données constitue l'un des fondements du développement d'une recherche en IA. Contrairement à une image trop souvent répandue, les législations française et européenne proposent un cadre suffisamment ouvert pour soutenir une recherche et une politique industrielle ambitieuses en la matière. Au-delà des possibilités évoquées ci-dessus, la création par le Règlement européen de protection des données (RGPD) d'un « droit à la portabilité », qui permet aux personnes de récupérer leurs données conservées par des acteurs privés, ouvre de grandes opportunités encore largement inconnues.

La puissance publique pourrait jouer un rôle moteur dans la concrétisation de ces dernières. Elle pourrait ainsi **lancer une grande cause nationale ou un grand projet de recherche fondé sur des données issues de la contribution de citoyens exerçant leur droit à la portabilité auprès des acteurs privés** et rebasculant leurs données pour un projet au service d'une cause d'intérêt général. L'Etat se porterait garant du respect des libertés par le projet et pourrait, par exemple, soutenir la mise en place d'un tableau de bord (sur le modèle du projet « NosSystèmes » de la FING) à la main des personnes. Au-delà de ce seul projet, la puissance publique amorcerait ainsi les potentialités ouvertes par la création du droit à la portabilité.

Les acteurs privés pourraient naturellement apporter leurs propres jeux de données à ce projet et participer à cette grande cause nationale.

## RECOMMANDATION 6

## Renforcer la fonction éthique au sein des entreprises

Identifier de possibles irrégularités ou effets néfastes en amont du déploiement d'algorithmes aux impacts significatifs, mais également assurer un rôle de veille en continu pour identifier les problèmes émergents, imperceptibles ou inaperçus au départ, en apportant un contrepoint à la perspective des opérationnels s'avère aujourd'hui une fonction essentielle des entreprises. Il s'agit aussi de développer une vision générale de chaînes algorithmiques dont on a souligné la tendance à la segmentation et à la compartimentation des fonctions et des préoccupations. Du même impératif participe la nécessité d'organiser des formes de dialogue entre opérationnels, personnalités extérieures à

64 Algorithmic Decision Making Systems.

65 Plusieurs projets ont déjà émergé de ce laboratoire, notamment, le projet de « dons-de-données » (en allemand « Datenspende Projekte » https://datenspende.algorithmwatch.org/), dans lequel plus de 4000 utilisateurs ont observé, pendant plusieurs mois avant et jusqu'à l'élection parlementaire allemande, les résultats de recherches Google concernant les 16 principaux candidats. L'idée sous-jacente de ce projet est de mesurer l'impact de la personnalisation par Google des résultats de recherches afin de confirmer ou d'infirmer la théorie appelée « filter bubble ».

66 https://www.nytimes.com/2017/11/14/opinion/academia-tech-algorithms.html

67 Des projets précédents ont montré que l'usage de données anonymisées était susceptible de générer des usages problématiques du point de vue éthique (ciblage de groupes de population – et non pas forcément d'individus – sur une base ethnique dans des contextes de conflit, segmentation actuarielle, etc.).

> Assurer un rôle de veille en continu pour identifier les problèmes émergents, imperceptibles ou inaperçus au départ s'avère aujourd'hui une fonction essentielle des entreprises

l'entreprise, acteurs et communautés impliquées par le fonctionnement des algorithmes ainsi que chercheurs en sciences humaines et sociales.

Plusieurs modalités de mise en œuvre de cet impératif pourraient être envisagées.

Une solution pourrait consister dans le déploiement de comités d'éthique au sein des entreprises déployant des algorithmes aux impacts significatifs. La composition et les modalités d'intervention de tels comités constituent un point essentiel. Publicité ou non des comptes rendus, publicité ou non de la composition du comité, degré éventuel d'indépendance: la palette des options possibles est large.

L'attribution de cet impératif à la fonction RSE ou aux déontologues pourrait également être envisagée.

Cette animation de la fonction de réflexion éthique dans le secteur privé pourrait aussi prendre la forme de réseaux constitués par secteurs ou branches professionnelles pour assurer la diffusion de bonnes pratiques ainsi que le repérage précoce de problèmes émergents. On pourrait d'ailleurs même considérer que des comités éthiques sectoriels puissent organiser une forme de veille éthique en lieu et place de comités installés au niveau de chaque entreprise, ce qui constituerait néanmoins une garantie moindre.

Ce travail en réseau devrait avoir pour objectif la constitution et la tenue à jour de référentiels éthiques sectoriels (chartes éthiques, codes de conduite, chartes de déontologie etc.), mais également la révision des codes d'éthique professionnels préexistants pour prendre en compte l'introduction des algorithmes et des systèmes d'IA.

Ces réflexions devraient en retour déboucher sur l'intégration, dans les chartes de déontologie des entreprises, d'un chapitre dédié aux enjeux soulevés par les algorithmes (en explicitant par exemple les limites à ne pas franchir en concevant les paramètres des systèmes, des obligations de qualité et d'actualisation des jeux de données utilisés pour entraîner les algorithmes, etc.).

Les diverses possibilités évoquées dans les paragraphes précédents ont pour but de souligner que la formule exacte à retenir mériterait sans doute de faire l'objet de débats spécifiques et que, à l'évidence, plusieurs positions peuvent exister.

# CONCLUSION

Les principes et les recommandations formulés à l'issue de ce rapport constituent le résultat de la synthèse par la CNIL des échanges et des réflexions menés à l'occasion du débat public national qu'elle a animé, grâce au soutien de soixante partenaires, de janvier à octobre 2017.

Les recommandations ont été formulées de façon très large, mobilisant tout le spectre possible des acteurs publics ou privés. Les défis soulevés par les algorithmes appellent une mobilisation, une attention et un questionnement de la part de l'ensemble des acteurs de la société civile (citoyens, entreprises, associations) pour piloter un monde complexe. Il ne s'agissait donc pas d'avancer que le véhicule à privilégier pour les appliquer ne pouvait être que la loi. Au contraire, la plupart des recommandations sont susceptibles d'être interprétées comme pouvant donner lieu, ou bien à une traduction juridique contraignante, ou bien à une appropriation volontaire de la part des acteurs, plusieurs degrés étant envisageables entre ces deux extrêmes.

Deux principes fondateurs ressortent de cette réflexion. Il convient d'y insister tout particulièrement, tant ils sont susceptibles de subsumer quelques-uns des défis éthiques majeurs soulevés par l'intelligence artificielle.

D'une part, **un principe substantiel, le principe de loyauté des algorithmes,** dans une formulation approfondissant celle déjà élaborée par le Conseil d'Etat (voir section « Le principe de loyauté »). Cette formulation intègre en effet une dimension de loyauté envers les utilisateurs, non pas seulement en tant que consommateurs, mais également en tant que citoyens, voire envers des collectifs, des communautés dont l'existence pourrait être affectée par des algorithmes, que ceux-ci d'ailleurs traitent des données personnelles ou pas.

D'autre part, **un principe plus méthodologique : le principe de vigilance.** Ce principe de vigilance doit être entendu, non comme une vague incantation mais comme une réponse étayée à trois enjeux centraux de la société numérique. Premièrement, le caractère évolutif et imprévisible des algorithmes à l'ère de l'apprentissage automatique (machine learning). Deuxièmement, le caractère très compartimenté des chaînes algorithmiques, induisant segmentation de l'action, indifférence aux impacts générés par le système algorithmique dans son ensemble, dilution des responsabilités. Troisièmement, enfin, le risque d'une confiance excessive accordée à la machine, jugée – sous l'effet d'une forme de biais cognitif humain – infaillible et exempte de biais. À travers le principe de vigilance, l'objectif poursuivi est en somme d'organiser l'état de veille permanente de nos sociétés à l'égard de ces objets socio-techniques complexes et mouvants que sont les algorithmes ou, à proprement parler, les systèmes ou chaînes algorithmiques. Un état de veille, autrement dit un questionnement, un doute méthodique. Ceci concerne au premier chef les individus qui composent les maillons des chaînes algorithmiques : il s'agit de leur donner les moyens d'être les veilleurs, lucides et actifs, toujours en questionnement, de cette société numérique. Ceci concerne aussi les autres forces vives de notre société. Les entreprises, bien sûr, pour modeler des systèmes algorithmiques vertueux, mais pas seulement.

Ces principes, par la démarche universelle dont ils procèdent, pourraient bien s'inscrire dans une nouvelle génération de principes et de droits de l'homme à l'ère numérique : cette génération qui après celles des droits-libertés, des droits patrimoniaux et des droits sociaux, serait celle des droits-système organisant la dimension sous-jacente à notre univers numérique. Ne sont-ils pas susceptibles d'être portés au niveau des principes généraux de gouvernance mondiale de l'infrastructure internet ? À l'heure où se construisent les positions française et européenne sur l'intelligence artificielle, la question mérite d'être posée.

**Les principes de loyauté et de vigilance pourraient s'inscrire dans une nouvelle génération de principes et de droits de l'homme à l'ère numérique : des droits-système organisant la dimension sous-jacente à notre univers numérique**

# ANNEXES

# Les applications et les promesses des algorithmes et de l'IA

Les pages qui suivent n'ont pas vocation à développer une vision critique. Il s'agit plutôt ici de décrire les grands usages des algorithmes et de l'IA déjà à l'œuvre ainsi que, dans un ordre plus prospectif, certaines promesses aujourd'hui évoquées principalement par des acteurs dont la posture n'est pas toujours neutre. Il convient de garder à l'esprit qu'une part non négligeable du discours public sur les algorithmes et l'IA – souvent la plus irénique et parfois la plus catastrophiste – est déterminée par des intérêts commerciaux.

## Santé

L'outil algorithmique fait l'objet de larges promesses. Comme dans chaque secteur, les opportunités dans le champ de la santé doivent cependant être appréhendées avec prudence, notamment du fait des immenses capacités « marketing » des organisations qui les déploient[68]. Le rôle annoncé et parfois déjà effectif des algorithmes et de l'IA dans le domaine de la santé est indissociable de l'existence de bases de données de plus en plus massives, tant en termes d'individus concernés qu'en terme de quantité de données disponibles sur chacun d'eux. **L'algorithme et l'IA permettent justement de tirer parti de cette quantité inédite de données disponibles aujourd'hui** (données issues des grandes bases médico-administratives rassemblées dans le SNDS[69] mais aussi des objets de santé connectée, des dossiers de patients, etc.) pour bâtir des modèles au sein desquels un profil très précis de chaque individu peut être dessiné, ce profil pouvant constituer le soubassement d'une prévision.

Ces promesses, qui concernent les grands objectifs de santé publique, concernent d'abord l'idée d'**une médecine à la fois prédictive, préventive et personnalisée.** L'analyse et la confrontation de mon profil génomique à celui d'individus similaires et à leurs parcours de santé peuvent aider au **diagnostic précoce.** Elles peuvent aussi permettre d'évaluer mes chances de développer telle ou telle maladie (cancer, diabète, asthme etc.), de « prédire » en quelque sorte ma santé future et dès lors m'inciter à prendre des mesures en conséquence (grâce à des campagnes de prévention ciblées). L'établissement de profils biologiques affinés permettrait également de **personnaliser les traitements et stratégies thérapeutiques.**

> **L'intelligence artificielle se développe notamment en cancérologie. L'un des exemples les plus fréquemment mentionnés à cet égard est celui de Watson, d'IBM. Watson analyse en effet les données génétiques des patients, les informations les concernant recueillies lors de leur admission, leur historique médical et les compare avec 20 millions de données issues d'études d'oncologie clinique pour établir un diagnostic et proposer un traitement. L'école de médecine de l'Université de Caroline du Nord a ainsi conduit en octobre 2016 une expérience montrant que les préconisations de Watson recoupaient les traitements prescrits par les cancérologues dans 99% des 1000 cas de cancer étudiés. Cette expérience a aussi démontré que dans 30% des cas, Watson était à même de proposer davantage d'options thérapeutiques que les médecins. Il convient toutefois de considérer avec prudence ces résultats annoncés, par ailleurs promus avec d'importants moyens de communication et de relations publiques.**

**68** Gérard Friedlander, doyen de la faculté de médecine de l'Université Paris Descartes, l'a notamment souligné (événement organisé par l'Hôpital Necker et l'Institut Imagine, le 15 septembre 2017)

**69** Système national des données de santé.

Outre la capacité d'une médecine de plus en plus appuyée sur les algorithmes et sur l'IA à faire fonds sur l'ensemble des variables biologiques, comportementales et environnementales, son intérêt réside aussi dans sa **capacité à traiter une masse d'informations scientifiques et de recherche qu'aucun médecin n'aurait matériellement la possibilité de maîtriser** (à titre d'exemple, on dénombre pas moins de 160 000 publications par an en cancérologie) dans la perspective de formuler un diagnostic.

L'intelligence artificielle est aussi susceptible de fournir un **appui à la détection de risques sanitaires. Les algorithmes peuvent être utiles pour « repérer l'élévation de l'incidence de maladies ou de comportements à risque, et d'alerter les autorités sanitaires[70] »**. Par exemple, en France, les liens entre l'utilisation d'une pilule contraceptive de 3ème génération et le risque d'AVC a pu être étudié grâce au traitement algorithmique de la base du Système national des données de santé (SNDS). La mise en œuvre d'un algorithme peut aussi permettre de **prédire des risques de maltraitance.** Il faut d'ailleurs souligner que ce n'est pas là un fait nouveau. Déjà en 1981, la CNIL avait eu à connaître d'un projet du Ministère de la Santé et des Affaires sociales visant à automatiser le signalement d'enfants présentant des risques psycho-sociaux (fichier GAMIN) : sur la base de l'analyse de 70 données différentes, le système devait repérer les cas à examiner en priorité. C'était bien un modèle, matérialisé par une série de critères, que l'algorithme permettait d'appliquer automatiquement à de très vastes cohortes.

Dans la pratique médicale, les algorithmes sont à certains égards déjà bien implantés pour **automatiser des tâches du quotidien.** Les logiciels d'aide à la prescription (LAP), une fois une maladie déjà diagnostiquée, sont déjà de précieux outils d'aide à la décision pour les médecins au moment de la saisie d'ordonnances. Ils permettent d'utiliser le dossier d'un patient pour repérer des contre-indications, des allergies ou des interactions médicamenteuses dangereuses. Autres applications déjà bien inscrites dans le paysage médical : « l'analyse d'image (imagerie médicale, anatomo-pathologie, dermatologie), l'analyse de signaux physiologiques (électro-cardiogramme, électro-encéphalogramme) ou biologiques (séquençage de génome) »[71]. L'utilité de l'IA est également mise aujourd'hui en avant pour **optimiser la mise en place d'essais cliniques grâce à une automatisation de la sélection des patients.**

En somme, sous réserve de précautions, l'algorithme en santé permettrait de mieux à répondre à certains besoins « pour les médecins (plus de sécurité), pour les patients (plus de personnalisation), et pour les instances publiques (plus de rationalisation) »[72].

## Éducation

L'application désormais la plus connue des algorithmes dans le domaine de l'éducation a trait à l'**affectation des immenses effectifs que doit gérer chaque année l'administration de l'Education nationale** et à l'attribution de places en lycée et dans le supérieur en fonction des vœux formulés par les candidats.

Le cas de l'algorithme « APB » (Admission post-bac) a été particulièrement évoqué depuis l'été 2016. Déployé depuis 2009 afin de faciliter et de fluidifier l'appariement entre les souhaits des élèves émis sous la forme de vœux et les places disponibles dans l'enseignement supérieur, il concernait en 2017 environ 808 000 inscrits dont 607 000 élèves de Terminale ayant candidaté aux 12 000 formations disponibles sur le logiciel. Pour que le système fonctionne, des critères de priorité ont été établis, pour être appliqués à tous. L'objectif poursuivi par APB, et sans évoquer à ce stade les critiques auxquelles il a pu donner lieu, étant double. D'une part, il s'agissait d'automatiser une tâche immense et donc d'optimiser un processus administratif particulièrement coûteux en temps. D'autre part, APB est aussi crédité d'avoir amélioré un fonctionnement qui laissait auparavant la place à des formes d'arbitraire. Roger-François Gauthier, expert des politiques éducatives, explique ainsi que les algorithmes tels qu'APB ou AFELNET (pour la répartition des élèves en lycée) « 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 ».

En plus d'être efficace, l'algorithme est donc présenté comme assurant l'équité et la non-manipulabilité (chaque lycéen devant effectuer ses choix sans autocensure) puisque la décision est prise de façon automatique, en fonction des données du candidat, de ses vœux d'affectation et selon des critères identiques pour tous puisque programmés une fois pour toutes lors du paramétrage de l'algorithme.

**L'autre grand champ d'application des algorithmes (y compris de l'IA) dans l'éducation concerne directement les pratiques pédagogiques elles-mêmes.** On s'y réfère souvent sous le titre de *learning analytics*. Ici encore le recours aux algorithmes est indissociable de la capacité inédite à collecter des données extrêmement nombreuses et diversifiées : données d'apprentissage (sur les résultats aux exercices mais potentiellement aussi sur la manière dont l'élève s'y confronte, durée de résolution), données sur les interactions avec l'enseignant et avec les pairs, données socio-démographiques, etc.

**70** INSERM, dossier d'information « Big data en santé » (disponible en ligne)
**71** Evénement organisé par le Conseil départemental du Rhône de l'Ordre des médecins, le 28 septembre 2017.
**72** Evénement organisé par le Conseil départemental du Rhône de l'Ordre des médecins, le 28 septembre 2017.

L'analyse des données sur l'apprentissage des élèves à l'aide d'algorithmes et de systèmes d'intelligence artificielle est aujourd'hui conçue comme le moyen de développer des stratégies de personnalisation de l'enseignement. On retrouve ici, comme dans le domaine médical abordé précédemment, la possibilité de la détermination d'un profil très fin de chaque élève mis à profit pour « diagnostiquer » une situation d'apprentissage, pour détecter un éventuel décrochage scolaire mais aussi pour permettre l'élaboration de stratégies individuelles d'apprentissage et de formation, adaptées au profil de chaque élève. La consultation en ligne organisée dans le cadre du débat public par le NumériLab (au sein du Ministère de l'Education Nationale) aborde les intérêts que pourraient revêtir les *learning analytics* pour la communauté éducative. Dans un contexte français où la différenciation pédagogique par « groupes de besoin » est loin d'être effective, certains contributeurs voient dans l'algorithme la possibilité de « mettre en œuvre des situations individuelles et collectives qui tiennent compte des différentes difficultés rencontrées par les élèves, leur permettre d'avoir des parcours individualisés, rendre compte de leur évolution et les faire partager à l'ensemble de la communauté éducative ».

## Vie de la cité et politique

Les algorithmes et l'intelligence artificielle investissent également le champ de la politique, au double sens du terme, c'est-à-dire tant en ce qui concerne l'organisation de la cité (les politiques publiques) que les pratiques de conquête du pouvoir, la vie électorale.

Le cas, précédemment évoqué, de l'algorithme APB en fournit un exemple. D'autres exemples sont peut-être moins attendus dans la mesure où il peut s'agir d'algorithmes déployés non pas par des administrations mais par des entreprises privées dont l'activité peut avoir un impact direct sur des domaines relevant généralement de l'autorité publique. Des applications de géolocalisation et de guidage routier mises à disposition des automobilistes peuvent modifier sensiblement les flux de circulation dans une ville et illustrent donc pleinement l'impact des algorithmes sur la vie collective. Dans un autre ordre d'idées, **la place cruciale que prennent désormais les algorithmes dans la recherche et le filtrage de l'information** les situent à un point névralgique de la vie démocratique. Les algorithmes de reconnaissance et le *machine learning* améliorent également l'efficacité de la modération automatique de propos déplacés sur les réseaux sociaux ou autres sites hébergeant des contenus : la DGMIC évoque l'ouverture par des sites de presse de leurs articles aux commentaires, l'algorithme permettant ici de favoriser le débat et l'exercice du pluralisme.

Au cours des dernières années, d'abord aux Etats-Unis, des offres de **logiciels d'aide à la stratégie électorale** se sont développées. Beaucoup de ces solutions reposent en fait sur la mise en œuvre d'algorithmes prédictifs qui analysent les données électorales. On retrouve ici, une fois encore, l'association entre la capacité des grandes quantités de données et celle – précisément par le biais de l'algorithme – à les exploiter, à les « faire parler » en construisant un modèle à partir de l'analyse des données passées qui est ensuite appliqué aux données actuelles pour élaborer enfin des recommandations, une aide à la décision stratégique.

Le logiciel 50+1 déployé depuis 2012 par la société LMP sur le fondement d'une expertise acquise pendant la campagne électorale américaine de 2008 en offre un bon exemple. Son objet est d'accompagner les stratégies électorales des candidats aux élections politiques en leur indiquant les zones à faire cibler en priorité par leurs équipes de militants pour des actions de porte-à-porte. L'algorithme intervient pour analyser les données des élections passées (résultats électoraux bureau de vote par bureau de vote, données socio-démographiques), en inférer un modèle qui, appliqué à la circonscription faisant l'objet de la campagne du candidat utilisateur du logiciel, permette in fine de formuler une prédiction sur la tendance dans l'aire de chaque bureau de vote. Outre l'intérêt que peut présenter ce type de logiciel pour les candidats aux élections, ses promoteurs le présentent également comme un moyen de lutter contre l'abstention dans le cadre d'une stratégie visant à remobiliser les abstentionnistes.

Si la législation française sur la protection des données personnelles ne permet pas le déploiement de logiciels qui cibleraient individuellement les électeurs (à l'exception de ceux y ayant consenti), le terrain électoral américain est un observatoire d'applications des algorithmes et de l'intelligence artificielle à des fins de profilage individuel[73]. Les deux campagnes présidentielles menées par Barack Obama ont vu se déployer de telles campagnes de marketing ciblé. La stratégie électorale de Donald Trump en 2016 semble avoir vu le franchissement d'un nouveau seuil dans le recours à ce type d'outils de communication ciblée[74], appuyés sur le recours à des données issues des réseaux sociaux et des courtiers en données. Même si de fortes incertitudes demeurent sur la réalité de ces pratiques, l'envoi de milliers de messages extrêmement individualisés (en fonction des préoccupations et attentes inférées du profil de chaque électeur) au cours d'une même soirée a pu être évoqué[75].

73 Une autre manière de présenter ces applications est de souligner qu'elles permettent une « prédiction » de ce que pourrait être le comportement d'un électeur ou encore une « recommandation » adressée à l'utilisateur du logiciel quant au type d'action requise en fonction du profil (envoi de tel message, utilisation préférentielle de tel canal de communication).

74 En France, les prédictions et recommandations d'un logiciel tel que 50+1 concernent des cohortes de 1000 personnes et ne relèvent donc pas d'outils de ciblage individuel.

75 https://www.theguardian.com/politics/2017/feb/26/robert-mercer-breitbart-war-on-media-steve-bannon-donald-trump-nigel-farage

## Culture et médias

L'utilisation d'algorithmes et d'intelligence artificielle produit déjà de forts impacts sur la structuration de l'offre de produits culturels et, partant, probablement aussi sur les pratiques de consommation culturelle. **Différents services de recommandation** facilitent la hiérarchisation de l'information « afin de répondre au besoin de l'utilisateur de s'orienter dans la surabondance des contenus accessibles »[76]. Le recours à ces services de « matching » concerne, à des degrés de développement variables selon la DGMIC, des secteurs très divers au sein de l'industrie culturelle : la vidéo à la demande par abonnement (80 % des contenus visionnés sur Netflix seraient issus de recommandations personnalisées), la musique (la fonctionnalité de recommandation apparaît comme un réel enjeu de différenciation pour des services de streaming tels que Spotify ou Deezer), les services gratuits (pour Facebook, Youtube ou encore Google, l'algorithme participe à la maximisation du nombre d'utilisateurs et ainsi à leur exposition augmentée à la publicité) ou encore les sites de commerces en ligne (30% des ventes d'Amazon résulteraient de ses recommandations algorithmiques).

L'intérêt de tels services est triple : tout d'abord, ils permettent de proposer au client une relation plus individualisée qui accompagne éventuellement la découverte d'autres offres. Ils peuvent également profiter à l'industrie audiovisuelle ou culturelle dans la mesure où ils facilitent *« la découverte d'œuvres audiovisuelles qui ne seraient pas par ailleurs programmées en raison de leur petit budget, ou en raison de l'absence d'un distributeur ou d'un budget de promotion ».* En effet, *« grâce aux moteurs de recommandation, certains films peuvent trouver un public même si ces films ne sont pas programmés par les chaînes de télévision traditionnelles »*[77]. Enfin – et peut-être, surtout – l'enjeu économique est indéniable pour fournisseurs et plateformes qui, en orientant les usagers, augmentent la satisfaction et ainsi l'utilisation de leur service.

Toutefois, il convient de nuancer les implications de l'algorithme sur le secteur. La DGMIC, qui s'appuie sur les auditions d'une quinzaine d'acteurs du secteur, indique que *« la recommandation personnalisée n'a pas tenu toutes ses promesses et que, bien que cette fonction apparaisse aujourd'hui incontournable à l'utilisateur, ce n'est pas celle qui guide majoritairement la consultation et la consommation des contenus »,* le travail des équipes éditoriales demeurant fondamental. En fonction du perfectionnement futur des algorithmes, la recommandation pourrait néanmoins occuper une place encore plus prépondérante dans les industries culturelles. A titre d'exemple, la startup Prizm commercialise des enceintes qui, en combinant des informations sur le moment de la journée, l'ambiance ou le nombre de personnes dans la pièce, diffusent la playlist la plus adéquate.

Ces algorithmes agissent sur le fondement de trois types de données : des données personnelles attachées aux profils des utilisateurs (l'historique d'usage par exemple), des données attachées aux œuvres (mots-clés indexés tantôt manuellement tantôt, depuis peu, de manière automatique) et, plus rarement, des données contextuelles (l'heure d'écoute ou la météorologie, par exemple). La recommandation peut s'exercer selon trois logiques différentes : tout d'abord, un filtrage sémantique peut viser à *« placer l'utilisateur sur la « cartographie » des **contenus »*** (DGMIC) sur la base notamment de son historique ou de questionnaires visant à mieux comprendre ses goûts. Une autre approche plus souvent privilégiée, celle du **filtrage collaboratif,** consiste à recommander en partant de l'hypothèse que deux utilisateurs partageant un avis sur un contenu sont plus susceptibles d'être également en accord sur un autre contenu plutôt que deux utilisateurs choisis aléatoirement. Le filtrage collaboratif peut se fonder autant sur le comportement des utilisateurs, leurs consommations passées *(« les utilisateurs ayant aimé le contenu A ont aussi aimé le contenu B »)* que sur l'objet directement *(« si Alice aime les contenus a, b, c et d, et que Benoit aime a, b et c, il est cohérent de recommander d à ce dernier »).* Enfin, un **filtrage hybride** permet de combiner ces deux méthodes pour optimiser la performance de recommandation. Ainsi, l'algorithme de recommandation de Spotify opère de la façon suivante[78] :

- En premier lieu, les contenus écoutés (style, tempo) sont analysés (soit plutôt une approche basée sur les contenus) ;
- Ensuite, les genres musicaux les plus appréciés sont regroupés en fonction de ceux des autres utilisateurs ayant consommé les mêmes contenus ;
- Enfin, l'utilisateur est circonscrit par rapport à ses propres comportements et son profil (type de consommation, fréquence d'écoute etc.).

Autre exemple d'algorithmes dans le champ de la culture : bien qu'embryonnaire, la **génération automatique de contenus culturels,** afin de produire et imiter les contenus qui plaisent, constitue une idée qui séduit. La DGMIC précise néanmoins que *« les nombreuses tentatives de prédiction du succès d'un livre ou d'un scénario déjà écrit grâce à des algorithmes n'ont pour l'instant pas permis de découvrir le secret le plus convoité des industries culturelles, industries de prototypes et de risques ».*

**76** DGMIC (Ministère de la Culture), « Les algorithmes dans les médias et les industries culturelles »
**77** Rapport du CSA Lab, p.14
**78** CNIL, « Les données, muses et frontières de la création », Cahier IP n°03, octobre 2015

## Justice

Des évolutions majeures sont à l'œuvre dans l'exercice des métiers du droit et de la justice. Si la plupart n'en sont qu'à un stade de développement anticipé, l'intérêt suscité par les outils technologiques recourant aux algorithmes est grand. L'algorithme peut tout d'abord permettre pour l'avocat ou le juge d'**obtenir un appui pour des tâches très variées désormais automatisables,** souvent répétitives voire laborieuses : « *l'évaluation des éléments de preuve selon différentes méthodes (fiabilité des témoins oculaires, distinction des rumeurs et des témoignages, procédures de discovery, constructions d'explications alternatives), la modélisation du travail des jurys, l'extraction d'informations contenues dans des documents (data mining), l'interprétation des informations (mise en lumière de modèles ou d'associations possibles, hiérarchisation des informations…), la recherche d'informations, la construction d'une argumentation (modélisation de structures d'argumentations, utilisation d'arbres de raisonnements permettant de lier une demande, aux justifications et objections dont elle peut faire l'objet), l'élaboration de documents, de formulaires juridiques et de contrats, ou encore la résolution de différends* »[79]. C'est cependant les promesses de la **justice dite « prédictive » ou « prévisionnelle »** qui méritent le plus d'attention tant elles pourraient bouleverser la conception « humaniste » de la justice. Elle peut être définie comme l'« outil informatique, reposant sur une base de données jurisprudentielles, qui, à l'aide d'algorithmes de tri et (pour les plus perfectionnés) de « réseaux neuronaux », va permettre d'anticiper quelles seront les statistiques de succès de tel ou tel argument juridique »[80].

La possibilité d'une justice prévisionnelle a été conditionnée par la présence de **bases de données jurisprudentielles** toujours plus fournies. Si elles n'étaient auparavant pas mises à la disposition de tous, la loi dite « République Numérique » a favorisé la diffusion des décisions de justice administrative et judiciaire par le mouvement d'ouverture des données publiques (open data) qu'elle consacre. De nombreuses start-ups[81] se saisissent ainsi depuis plusieurs mois de cette nouvelle masse de données pour développer leurs outils de « justice prévisionnelle » dont l'objectif principal est de repérer des récurrences à des fins de prédiction. Les intérêts et exploitations potentiels sont multiformes et recouvre différents champs des métiers du droit, du juriste à l'avocat en passant par le juge.

Pour le **justiciable et les professionnels du droit,** les logiciels algorithmiques peuvent être d'une utilité stratégique en optimisant l'**identification des solutions statistiquement les plus probables pour un contentieux donné** ou le montant prévisible des dommages-intérêts. Cette méthode est notamment adaptée aux contentieux particulièrement propices aux récurrences, tels que le licenciement sans cause réelle et sérieuse ou les prestations compensatoires en cas de divorce. Si ces outils se généralisent au sein des différentes professions du droit, ils contribueraient au dessein plus général d'une « smartjustice », à savoir une justice animée par des impératifs de meilleure rentabilité avec le minimum de moyens, grâce aux technologies. L'avocat pourrait bénéficier d'un gain significatif de temps – et ainsi se consacrer à des tâches plus gratifiantes d'analyse juridique et de contact humain –, tandis que le justiciable pourrait éviter certains coûts, en faisant le choix de s'entendre à l'amiable plutôt que de saisir le juge dans des cas où les chances du succès d'un procès sont réduites. Pour le fonctionnement du système de justice français dans son ensemble de surcroît, un recours grandissant à ces solutions annoncerait une diminution du nombre de saisines et un certain désengorgement des juridictions, éventualité plus qu'attrayante dans le contexte actuel.

Parallèlement, les algorithmes prédictifs peuvent être une **ressource utile au juge :** s'inspirer des recommandations de la « machine », fondées sur les jurisprudences précédentes, lui permettrait d'éclairer ses décisions. C'est l'ambition d'une qualité de la justice augmentée et de l'**harmonisation des décisions** qui serait au cœur de ce modèle. En d'autres termes, dans la continuité de mesures récentes telles que les barèmes, la « justice prévisionnelle » réduirait l'horizon d'incertitude et participerait à l'évaluation interne des juridictions et magistrats.

**L'aide à la décision fondée sur des algorithmes au service du juge a fait l'objet d'une expérimentation cette année par les cours d'appel de Rennes et Douai en partenariat avec le ministère de la Justice. Par un communiqué du 9 octobre 2017, le ministère a annoncé que l'outil développé par la start-up Predictice ne s'était pas avéré satisfaisant. De futures expérimentations sont néanmoins prévues, aux prochains stades du développement de l'outil.**

Le mouvement d'open data n'en est qu'à ses premiers frémissements : toutes les conditions sont ainsi réunies pour que la justice prévisionnelle prolifère, dans un contexte où 1,5 million de décisions seront désormais « anonymisables » chaque année et ainsi mises à disposition sur Jurinet (base interne de la Cour de cassation) et Légifrance.

79 Contribution au débat public d'un groupe de travail du Conseil National des Barreaux.
80 BOUCQ Romain, « La justice prédictive en question », Dalloz Actualité, 14 juin 2017.
81 Predictice, Case Law Analytics ou encore Doctrine.fr constituent des exemples de telles *legaltech*.

## Banque, Finance

Plusieurs événements récents ont contribué à accroître l'attention publique accordée à l'utilisation d'algorithmes dans le champ financier. Même si le rôle effectif joué par l'algorithme dans le flash crash du 6 mai 2010 pose débat, la chute historique du Dow Jones (environ 9%) a largement interpellé quant aux risques d'emballement, de manipulation et de comportements moutonniers associés à ce qui est aujourd'hui désigné sous l'appellation **« trading haute fréquence »** (THF). Un autre exemple est celui du piratage du compte Twitter de l'agence Associated Press en avril 2013 : en surveillant les mots-clefs figurant sur le réseau social, les algorithmes ont conclu à un attentat à la Maison blanche, précipitant ainsi le retrait de milliards d'ordres sur les marchés en quelques secondes.

Aussi appelé *speed trading* ou trading algorithmique, le THF désigne l'**automatisation d'arbitrages boursiers** qui connaît une ascension fulgurante depuis la fin des années 1990. La directive dite « MIF II »[82], qui encadre cette tendance et qui entrera pleinement en application en janvier 2018, définit le THF comme *« la négociation d'instruments financiers dans laquelle un algorithme informatique détermine automatiquement les différents paramètres des ordres [...] avec une intervention humaine limitée ou sans intervention humaine »* (article 4, paragraphe 39). Ce négoce financier d'un genre nouveau voit ainsi des robots traders d'une rapidité remarquable se substituer aux traditionnels « teneurs de marchés ». Prendre des décisions d'investissement et organiser les liquidités ne constituent désormais plus l'apanage de l'humain : le robot serait impliqué dans près de 70 % des transactions aux Etats-Unis et environ 40 % de celles en Europe. L'algorithme jalonne désormais le processus d'investissement, d'abord en amont par l'identification des opportunités, puis en aval par les règles opératoires d'exécution qui prennent position à l'achat ou à la vente[83]. Peu de problématiques inédites semblent finalement avoir émergé depuis une quinzaine d'années, et le THF a déjà fait l'objet d'un important effort de régulation et de responsabilisation des acteurs[84].

Si un tel intérêt pour l'automatisation s'est manifesté depuis 1995, année de création du THF, c'est parce qu'il redéfinit radicalement la temporalité de la bourse : dans ce secteur, l'ampleur de l'écart entre performance humaine et performance technologique est incontestable. Cette vélocité implique un second intérêt de taille pour les acteurs du THF (gestionnaires de fonds, institutions bancaires)[85] qui, en mettant en œuvre des stratégies d'investissement irréalisables manuellement, voient leurs profits et leur compétitivité croître de manière substantielle.

L'algorithme dans le champ de la finance, c'est également l'émergence de « robo-advisors » visant à **automatiser les services financiers** fournis aux clients et la gestion de leurs portefeuilles. Leur niveau de maturité semble néanmoins à ce stade assez faible selon l'Autorité des marchés financiers. Plus facilement mobilisables aujourd'hui, des outils basés sur des algorithmes visent à **automatiser la gestion des risques** et le contrôle de la conformité (la lutte anti-blanchiment, par exemple). Enfin, la personnalisation permise par l'algorithme pourrait mener à des mutations importantes des services financiers proposés aux clients : et si, à terme, une segmentation s'opérait entre une clientèle réduite qui aurait accès à des produits très sophistiqués comparativement à une autre qui se verrait proposer un éventail très réduit de produits simples ?

## Sécurité, défense

Le recours aux algorithmes dans le domaine de la sécurité et de la défense a pour finalités principales annoncées l'identification de suspects, la prédiction de commission d'infractions et l'automatisation d'opérations de maintien de l'ordre voire de guerre, jusqu'à l'acte de tuer, cette dernière finalité faisant l'objet d'un vif débat international.

Les années 2000 ont vu converger accroissement de la menace terroriste dans le sillage du 11 septembre 2001 et de l'explosion du nombre de données disponibles (liée à la numérisation globale des sociétés). Le problème, classique pour le renseignement, de l'analyse et de l'exploitation des données disponibles s'en trouve accru. Dans un contexte d'exigence politique très forte à l'égard des services de renseignement, les algorithmes sont présentés – à tort ou à raison – comme une solution permettant d'identifier les suspects en tirant pleinement partie des données disponibles. Le système API-PNR ou les « boîtes noires » évoquées lors de la discussion de la loi sur le renseignement de 2015 relèvent de cette logique. Les données des passagers aériens, dans un cas, celles de l'ensemble de la population, dans l'autre, sont filtrées par des algorithmes à la recherche de « signaux faibles », de profils considérés comme suspects en fonction de critères tenus secrets.

L'essor de l'idée de **« police prédictive »** correspond à l'idée de prédire la commission d'infractions au moyen de l'analyse massive de données concernant la commission passée de crimes et de délits afin de répartir plus efficacement les patrouilles. La promesse portée par exemple par le logiciel américain « Predpol » est d'adosser les méthodes policières traditionnelles jugées trop subjectives par des méthodes considérées comme « objectives ».

**82** Directive 2014/65/UE du Parlement européen et du Conseil du 15 mai 2014 concernant les marchés d'instruments financiers.
**83** Benghozi P.-J., Bergadaà M., Gueroui F., Les temporalités du web, 2014, chapitre 3 «Trading haute fréquence : l'arbitre sans sifflet ».
**84** Directive 2014/65/UE du Parlement européen et du Conseil du 15 mai 2014 ; loi n° 2013-672 du 26 juillet 2013 de séparation et de régulation des activités bancaires.
**85** Getco, Flow traders, IMC, Quantlab, Optiver...

Ces initiatives soulèvent pourtant d'importantes critiques. Le sociologue Bilel Benbouzid conteste ainsi à propos du logiciel « PredPol » la pertinence de l'application à la criminalité de logiques empruntées à la sismologie. L'attrait présenté par « Predpol » aux yeux de décideurs tiendrait en revanche à ce que ce type d'outils permet de *« gérer, selon des critères gestionnaires, l'offre publique de vigilance quotidienne »*[86]. Par ailleurs, leurs résultats concrets semblent pour l'heure décevants aux praticiens eux-mêmes. Par exemple, l'expérimentation « PredVol » visant la prédiction des actes de délinquance commis sur les véhicules aboutit « à faire ressortir toujours les mêmes spots, les mêmes points chauds aux mêmes endroits » selon le Colonel Philippe Mirabaud. Aussi, l'étude des vols avec violence sans arme contre les femmes sur la voie publique à Paris[87] révèle la forte régularité de ces actes – aussi bien en termes de localisation que d'horaires – mais les possibilités prédictives restent limitées. Ce constat repose en partie sur les difficultés à « faire parler » des jeux de données encore disparates. La plupart des outils dits « prédictifs » s'appuient sur les données des préfectures de police – plaintes des victimes et/ou arrestations notamment – dont l'utilisation à des fins de prédiction est loin d'être naturelle. Le perfectionnement de ces logiciels implique de les compléter par des données externes concernant autant le terrain des actes (densité de bars ou commerces, présence d'une station de métro etc.) que les conditions météorologiques ou encore les événements organisés au sein d'une ville par exemple.

Les experts invitent à ne pas sombrer dans le fétichisme technologique en pensant que l'algorithme aurait la capacité d'apporter une solution magique aux enjeux de sécurité. Gilles Dowek explique ainsi que « même avec un système d'une performance extrêmement élevée, il y aura toujours beaucoup plus d'innocents que de coupables accusés. Supposons un algorithme d'une super-qualité qui n'a qu'une chance sur 100 de se tromper. Sur 60 millions de personnes, ça fait 600 000 personnes détectées à tort, plus les 1 000 « vrais positifs » qu'on a bien détectés. Donc l'algorithme détecte 601 000 personnes, parmi lesquelles en réalité 1 000 seulement sont de vrais terroristes ». Le risque est donc de démultiplier la suspicion et de confronter les services de renseignement à une masse de cibles impossible à traiter.

Au-delà de l'identification de zones à risque, l'algorithme peut servir d'**aide à la résolution des enquêtes.** En s'appuyant sur les connexions entre l'ensemble des pièces d'une enquête – dont celles au caractère très technique comme procès-verbaux, appels téléphoniques ou encore informations bancaires –, des logiciels offriraient aux gendarmes la possibilité d'identifier des relations que l'humain n'était jusqu'ici pas parvenu à effectuer[89].

Si tout invite à modérer l'enthousiasme des promoteurs de la police prédictive, des perspectives intéressantes résident dans l'appui à la contextualisation, à l'interprétation et ainsi à l'organisation. Hunchlab, projet de l'entreprise Azavea, œuvre ainsi en ce sens en accentuant l'effort d'intelligibilité quant à ce que la prédiction permet ou ne permet pas, par une rétroaction plus solide et une interaction plus forte entre l'humain et l'outil. Dans tous les cas, s'il est difficile de préciser les contours des futures solutions privilégiées et malgré des réserves au sein de la communauté scientifique, les pouvoirs publics n'excluent pas d'y avoir recours pour *« la constitution d'une aide à la décision (« analyse décisionnelle »), au profit du commandant d'unité territoriale, notamment à des fins de prévention de la délinquance »*[90].

Extrêmement sensible est enfin la question posée par le **développement d'armes létales autonomes (robots tueurs)** qui pourraient prendre elles-mêmes la décision de tuer sur le champ de bataille ou à des fins de maintien de l'ordre. De tels systèmes sont déjà déployés à la frontière entre les deux Corée et les armées de divers pays réfléchissent actuellement à la mise en service de drones tueurs capables d'engager et d'éliminer une cible sans intervention humaine. En 2015, une pétition signée par plus d'un millier de personnalités, dont une majorité de chercheurs en IA et en robotique, ont réclamé l'interdiction des armes autonomes, capables *« de sélectionner et de combattre des cibles sans intervention humaine »*. Cette initiative a donné de la visibilité à un débat international déjà engagé à l'ONU.

## Assurance

Les algorithmes offrent tout d'abord au secteur assurantiel la possibilité d'accélérer et de **fluidifier des pratiques quotidiennes** telles que la gestion des sinistres, le suivi du comportement des assurés, leur indemnisation ou encore la lutte contre la fraude. La reconnaissance d'images permise par l'IA pourrait mener à systématiser, grâce à l'analyse des images de sinistres, les processus d'indemnisation « auto » et habitation . Autre exemple : l'intelligence artificielle, en révélant des liens insoupçonnés, peut se révéler utile pour retrouver les titulaires de contrats d'assurances-vie en déshérence (non réclamés) ou leurs héritiers. Mais, plus qu'un instrument d'automatisation au service de pratiques déjà bien implantées, l'algorithme annonce un nouveau paradigme de l'assurance et du mutualisme au sens où il pourrait *« modifier la manière d'appréhender les risques et de les valoriser, transformer les techniques et les pratiques de mutualisation »* (François Ewald[92]).

Certes la donnée a toujours constitué la matière première pour l'assureur pour prévenir les risques. Assurer, c'est pro-

**86** Bilel Benbouzid, « A qui profite le crime ? », *La Vie des Idées.* « PredPol » prétend s'inspirer des méthodes de prédiction des tremblements de terre pour offrir une « analyse du crime en temps réel qui prend la forme d'un tableau de bord ».
**87** Le géostaticien Jean-Luc Besson l'a exposée lors du même événement organisé par l'INHESJ.
**88** http://tempsreel.nouvelobs.com/rue89/rue89-internet/20150415.RUE8669/l-algorithme-du-gouvernement-sera-intrusif-et-inefficace-on-vous-le-prouve.html
**89** Le loigiciel AnaCrime a ainsi permis il y a quelques mois de relancer l' « affaire Gregory ».
**90** Réponse du Ministère de l'Intérieur à la question n°16562, publiée dans le JO Sénat du 29 décembre 2016.
**91** Evénement organisé par la Fédération Française de l'Assurance, le 5 juillet 2017.
**92** François Ewald, « After Risk, vers un nouveau paradigme de l'assurance. L'Assurance à l'âge du Big Data », septembre 2013.

poser des services financiers de protection différents selon des profils de risque, grâce au traitement de données à caractère prédictif. Florence Picard (Institut des actuaires)[93] explique ainsi comment l'actuaire a toujours eu pour rôle de de calculer la probabilité et l'impact financier des risques. Loin de l'appréciation globale fondée sur les déclarations des clients, l'algorithme annonce une évaluation plus fine s'appuyant sur des données comportementales en masse.

Objets connectés, réseaux sociaux, données de santé : de nouveaux horizons s'ouvrent vers une personnalisation sans précédent. Ce sont des corrélations inédites et individualisées qui sont ici recherchées. A titre d'exemple, certains assureurs auraient constaté que les clients achetant des feutres à placer sous les pieds de table et de chaise, pour la préservation du bon état de leur parquet, ont un comportement automobile bien plus prudent que la moyenne, et qu'une réduction de prime apparaîtrait ainsi comme justifiée[94]. Le risque peut dès lors être bien plus subjectivisé, individualisé. C'est, selon François Ewald, le passage de la notion de risque comme événement à celle de risque de comportement : *« les risques [...] étaient d'abord appréhendés par leurs caractéristiques objectives, à partir des événements qui en marquent la réalisation [...] on peut désormais les observer comme caractéristiques du comportement des agents »*. L'**assurance comportementale** concerne déjà les comportements des automobilistes à travers les coefficients de réduction-majoration (plus connus sous le nom de bonus ou malus) fondement du « pay as you drive » (qui se fonde sur les antécédents des conducteurs notamment en termes de vitesse moyenne). Plus encore, l'adaptation du coût des offres selon le style de conduite (accélérations, freinages brusques) est également possible grâce aux capteurs dont certains véhicules sont équipés *(« pay how you drive »)*.

En santé, l'individualisation ne mène pas encore à une segmentation tarifaire explicite, mais le programme Vitality de la société Generali révèle comment un système de récompenses peut indirectement permettre de s'adapter aux comportements des clients. Vitality se présente comme un programme visant à améliorer le bien-être en mettant à disposition *« des recommandations et des outils pour [...] encourager à mener une vie plus saine »* et en récompensant ceux qui atteignent leurs objectifs – plutôt qu'en pénalisant ceux qui ne les atteignent pas – *« grâce à des réductions et des offres avantageuses »* chez des partenaires[95]. *« Likes »* émis sur les réseaux sociaux et données de profil sur Facebook, par exemple, pourraient également servir à proposer des tarifs automobile avantageux.

Qui dit risque individualisé dit possibilité de **services de prévention augmentés** qui vont *« agir directement sur la source du risque pour tenter d'éviter qu'il survienne »* (par des incitations relatives à l'activité physique, la nutrition etc.), comme l'explique Fabrice Faivre (MACIF)[96]. Argument de santé publique en faveur des assurés, l'individualisation annonce aussi des modèles renégociés en assurance, fondés sur la détection des profils à haut risque et sur une nouvelle relation possible des assureurs au client.

La portée de cette personnalisation mérite toutefois d'être nuancée : une segmentation tarifaire trop fine ne serait pas nécessairement dans l'intérêt d'un assureur, du fait des conséquences potentielles qu'impliquerait une erreur en absence de volumes significatifs. Il reste encore à savoir si le cadre légal actuel est adapté à un tel mouvement. Celui-ci tend à restreindre l'utilisation de données par les assureurs au nom de principes tels que la protection des données à caractère personnel (le SNDS créé par la loi de « modernisation de notre système de santé » limite l'accès des assureurs aux données de santé) ou la lutte contre les discriminations. A titre d'exemple, la Cour de justice de l'Union européenne interdit l'utilisation du genre comme variable au sein d'un modèle statistique en assurance au nom du principe d'égalité de traitement entre les hommes et les femmes[97]. Florence Picard (Institut des Actuaires)[98] émet des réserves quant à la pertinence d'écarter un tel critère qui pourrait trouver toute sa place dans certains modèles. Cécile Wendling (AXA)[99] mentionne comment cette obligation du droit européen pourrait toutefois être mise à mal si les algorithmes de *machine learning* venaient à se déployer, inférant ainsi par eux-mêmes les critères permettant d'identifier le genre (couleur d'un véhicule etc.). En somme, l'avenir de ces pratiques semble aujourd'hui dépendre, d'une part, du degré de développement futur des objets connectés et, d'autre part, de la volonté qu'auront les citoyens de transmettre volontairement et consentir à l'utilisation de certaines des données les concernant pour améliorer leur santé par exemple.

## Emploi, RH, recrutement

Alors que chômage et conditions de travail sont au centre des préoccupations sociétales, les initiatives convoquant des algorithmes foisonnent pour répondre aux grands enjeux du marché de l'emploi.

Les particuliers, d'une part, peuvent recourir à des **agrégateurs d'offres d'emploi** qui se perfectionnent. L'APEC (Association pour l'emploi des cadres) dispose par exemple d'un algorithme sémantique permettant non seulement de rechercher les offres d'emploi selon des mots-clés, mais également d'induire automatiquement d'un CV le référentiel de compétences et de talents qui permettra de suggérer ensuite les offres d'emploi le plus finement possible[100]. Pôle

93 Evénement organisé par la Ligue des Droits de l'Homme, le 15 septembre 2017.
94 Dominique Cardon, *A quoi rêvent les algorithmes,* Paris, 2015, p.52
95 Site officiel de Vitali.
96 Evénement organisé par la Ligue des Droits de l'Homme, le 15 septembre 2017.
97 CJUE, 1er mars 2011, affaire C-236/09 dite « Test-Achats »
98 Evénement organisé par Fotonower, le 22 septembre 2017.
99 Evénement organisé par la Chaire IoT de l'ESCP Europe, le 20 septembre 2017.
100 Le Directeur des systèmes d'information de l'APEC a présenté cet outil lors de l'événement organisé par FO-Cadres, le 18 avril 2017.

Emploi dispose également de sa propre solution algorithmique d'agrégation des offres d'emploi.
Dans le cadre d'une réflexion éthique, c'est surtout l'appropriation par les entreprises – et notamment les directions des ressources humaines (DRH) – qui interpelle. Sans impliquer une réelle rupture des missions traditionnelles du DRH, l'algorithme faciliterait l'atteinte des objectifs de **recrutement** et de **gestion des ressources humaines,** à savoir : répondre aux attentes de rapidité dans la mobilité et le recrutement, accentuer l'emprise des collaborateurs sur leur propre parcours et, enfin, mettre à la disposition des managers la ressource adéquate pour l'atteinte des objectifs[101]. C'est notamment le coût de la « mauvaise embauche » qui pourrait être évité, voire la réduction à une échelle plus large du chômage. Prédire pour mieux satisfaire collaborateur et manager, en tirant profit d'une masse de données pertinentes : telle est la promesse de l'algorithme.

**Recrutement :** L'algorithme peut constituer un **instrument de « matching » affinitaire,** basé sur la sémantique, mobilisé par le DRH pour préqualifier le volume parfois conséquent de CV reçus pour une offre d'emploi donnée. Au vu de la durée moyenne très réduite de qualification d'un CV par un humain, nombreux sont ceux qui invoquent la plus grande « rigueur » de l'algorithme. La Harvard Business Review publiait ainsi une étude en 2014 affirmant qu'un algorithme peut surpasser le recruteur humain et éviter des présupposés fréquents chez ce dernier, tels que la tendance à corréler systématiquement prestige d'une université et performance future.
**Gestion des RH :** L'algorithme pourrait identifier les collaborateurs les plus à même d'être performants dans un rôle déterminé, en allant chercher ceux qui n'auraient pas candidaté à une offre de poste donnée. La mobilité interne serait également optimisée par le ciblage de formations appropriées pour un collaborateur afin d'esquisser pour lui un nouveau cheminement de carrière.
**Qualité de vie :** D'autres applications, pour lesquelles la réflexion éthique est plus que de rigueur, concernent la compréhension de certains phénomènes sociaux au sein de l'entreprise : analyse des facteurs justifiant l'absentéisme, prédiction des risques psychosociaux, calcul du risque de départ d'une organisation etc.

C'est bien la donnée qui constitue ici le matériau à faire fructifier pour résoudre certains défis du marché de l'emploi, à commencer par l'insuffisance de méthodes traditionnelles de recrutement (CV, entretiens) considérées lacunaires pour « atteindre » l'intimité de l'individu. Aptitudes comportementales et cognitives (« soft skills ») sont désormais plus activement recherchées : l'algorithme capitalise sur la donnée répartie au sein voire à l'extérieur de l'entreprise pour œuvrer en ce sens.

Ce sont d'abord les **données internes** à une organisation qui peuvent alimenter l'algorithme, à condition qu'elles puissent être rassemblées et fiabilisées. Si ces données existent, elles sont *« détenues pour partie par les collaborateurs, pour partie par les fonctions RH et pour partie pour les managers »*[102] et elles s'inscrivent plus dans une *« logique de "rendre compte" que d'exploitation pour des motifs précis »*[103].

Parfois jugées insuffisantes, la collecte de **données externes** (telles que les informations relatives aux parcours professionnels publiées sur les réseaux sociaux) est également attrayante pour de nombreuses organisations.

Les traitements algorithmiques semblent encore aujourd'hui limités[104], les quelques exemples existants aujourd'hui ne recourant pas à l'intelligence artificielle et au *machine learning*. Le frémissement est cependant tangible : le nombre de start-ups ayant pour objet les RH est passé de 200 à 600 en l'espace de deux ans[105]. Il est trop tôt pour évaluer de manière certaine l'impact qu'auront ces technologies sur les pratiques de recrutement et la gestion des talents[106]. N'est-il pas, par exemple, trop ambitieux d'affirmer qu'un algorithme pourrait se substituer à l'homme pour effectuer le travail conséquent qu'implique l'évaluation annuelle des collaborateurs ? En matière de recrutement, l'abandon par Google de son système de tri automatisé des candidatures semble révéler certaines limites de l'algorithme. Le perfectionnement des outils pourrait émaner de nouveaux développements en intelligence artificielle – des algorithmes qui construiraient leurs propres référentiels – ou encore d'un recours grandissant à la robotique (l'analyse de l'expressivité émotionnelle lors d'un entretien de recrutement en constitue une illustration[107]). Ces nouveaux outils pourraient engendrer de nouveaux risques : comment distinguer en RH les décisions simples facilement automatisables des décisions complexes pour lesquelles la dimension « humaine » de la profession devra être préservée ?

101 Analyse de Jean-Cristophe Sciberras, directeur des RH France et directeur des relations sociales corporate chez Solvay, lors de l'événement organisé par la CFE-CGC.
102 Sabine Frantz lors de l'événement de FO-Cadres.
103 Béatrice Ravache lors de l'événement de la CFE-CGC.
104 Le service des « questions sociales et RH » de la CNIL ne recense aujourd'hui aucune demande d'autorisation pour de tels traitements à dimension prédictive et n'est consulté qu'à titre informationnel afin de connaître l'avis de la Commission.
105 Jérémy Lamri, fondateur du LabRH, lors de l'événement de la CFE-CGC.
106 Béatrice Ravache, lors de l'événement de la CFE-CGC, évoque au sujet de la fonction de DRH que *« le savoir n'est plus la mémoire, n'est plus l'organisation des données, ni aller les chercher, c'est encore autre chose qui n'est pas forcément évident pour le RH aujourd'hui ».*
107 Laurence Devillers, événement CFE-CGC.

# REMERCIEMENTS

La CNIL adresse ses plus vifs remerciements aux personnes et aux institutions qui ont apporté leur participation à cette réflexion collective.

## Les partenaires du débat public

- Académie des technologies
- Agence Française de Développement (AFD)
- Association française de droit du travail et de la sécurité sociale (AFDTSS)
- Association française pour l'intelligence artificielle (AFIA)
- Caisse des dépôts et consignations (CDC)
- Centre de recherche de l'école des officiers de la gendarmerie nationale (CREOGN)
- Collège des Bernardins
- Comité consultatif national d'éthique (CCNE)
- Comité d'éthique du CNRS (COMETS)
- Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique (CERNA) d'Allistene
- Communication Publique
- Confédération française de l'encadrement – Confédération générale des cadres (CFE-CGC)
- Conseil départemental du Rhône de l'Ordre des Médecins
- Conseil National des Barreaux (CNB)
- Conseil Supérieur de l'Audiovisuel (CSA)
- Conservatoire National des Arts et Métiers (CNAM)
- Cour administrative d'appel de Lyon
- Cour d'appel de Douai
- École des Hautes Etudes en Sciences Sociales (EHESS)
- École Nationale Supérieure de Cognitique (ENSC)
- ESCP Europe, Chaire IoT
- Etalab
- Faculté de Droit de l'Université Catholique de Lille, Centre de recherche sur les relations entre le risque et le droit
- Faculté de Droit de l'Université Catholique de Lyon
- Familles rurales
- Fédération Française de l'Assurance (FFA)
- FO-Cadres
- Fondation Internet Nouvelle Génération (FING)
- Fotonower
- Génotoul societal
- Groupe VYV (MGEN – ISTYA – Harmonie)
- Hôpital Necker
- INNOvation Ouverte par Ordinateur (INNOOO)
- Institut des Hautes Etudes de Défense Nationale (IHEDN)
- Institut des Systèmes Complexes de Paris Ile-de-France (ISC-PIF)
- Institut Imagine
- Institut Mines-Télécom (IMT), Chaire de recherche Valeurs et Politiques des Informations Personnelles
- Institut National des Hautes études de la Sécurité et de la Justice (INHESJ)
- Institut National des Sciences Appliquées (INSA)
- Laboratoire pour l'Intelligence Collective et Artificielle (LICA)
- Le Club des Juristes
- Ligue de l'Enseignement
- Ligue des Droits de l'Homme (LDH)
- Microsoft
- Ministère de l'éducation nationale, via la direction du numérique pour l'éducation (DNE) et son Numéri'lab
- Ministère de la Culture, via la direction générale des médias et des industries culturelles (DGMIC)
- OpenLaw
- Ordre des avocats de Lille
- Randstad
- Renaissance Numérique
- Sciences Po Lille
- Sciences Po Paris
- Société informatique de France (SIF)
- The Future Society at Harvard Kennedy School, AI Initiative
- Universcience
- Université de Bordeaux
- Université de Lille 2
- Université Fédérale de Toulouse
- Université Paris II
- Visions d'Europe

## Les autres contributeurs

- Arbre des connaissances
- Autorité de contrôle prudentiel et de résolution (ACPR)
- Autorité des marchés financiers (AMF)
- Montpellier Méditerranée Métropole et son président, M. Philippe Saurel
- Ville de Montpellier

Les 37 citoyens ayant pris part à la concertation citoyenne organisée à Montpellier le 14 octobre 2017.

Jérôme BERANGER • Nozha BOUJEMAA • Dominique CARDON • Jean-Philippe DESBIOLLES • Paul DUAN • Flora FISCHER • Antoine GARAPON • Roger-François GAUTHIER • Hubert GUILLAUD • Rand HINDI • Jacques LUCAS • Camille PALOQUE-BERGES • Bruno PATINO • Antoinette ROUVROY • Cécile WENDLING

# LISTE DES MANIFESTATIONS ORGANISÉES **DANS LE CADRE DU DÉBAT PUBLIC**

De fin mars à début octobre, la CNIL a assuré l'animation et la coordination de 45 événements sur les algorithmes et l'intelligence artificielle. Certaines initiatives ont été imaginées spécifiquement à l'occasion du lancement du débat public, d'autres s'inscrivaient déjà dans les projets d'acteurs – institutions publiques, associations, centres de recherche – déjà préoccupés par ces enjeux.

De nombreux acteurs ont fait le choix d'appréhender les algorithmes dans un secteur spécifique (santé, emploi ou éducation par exemple) alors que d'autres ont abordé l'objet technologique dans sa globalité. Enfin, ce sont autant des ateliers d'experts à public restreint que des manifestations orientées vers l'appropriation du grand public (citoyens, étudiants etc.) qui ont jalonné ce processus.

**Plus d'informations sur ces manifestations sont disponibles sur le site de la CNIL.**

23/01/2017 — EVÉNEMENT DE LANCEMENT :
**TABLES-RONDES** « Des algorithmes et des hommes »
et « Loyauté, transparence et pluralité des algorithmes »
**> CNIL**

23/03/2017
25/03/2017 — **COLLOQUE** « Vers de nouvelles humanités ? »
**> Universcience**

31/03/2017 — **CONFÉRENCE** « Les algorithmes et le droit »
**> Université de Lille II**

06/04/2017 — **CONFÉRENCE** « Le choix à l'heure du Big Data »
**> Sciences Po Lille et Visions d'Europe**

08/04/2017 — **DÉBAT** « The governance of emerging technosciences »
**> German American Conference at Harvard University**

18/04/2017 — **DÉBAT** « Transatlantic perspectives on: AI in the age of social media;
privacy, security and the future of political campaigning »
**> The Future Society at Harvard Kennedy School**

18/04/2017 — **TABLES-RONDES** « Big Data, ressources humaines : les algorithmes en débat »
**> FO-Cadres**

04/05/2017 — **CONFÉRENCE** « Loyauté des décisions algorithmiques »
**> Université Toulouse III – Paul Sabatier**

16/05/2017 — **DÉBAT** « Le numérique tuera-t-il l'Etat de droit ? »
**> Collège des Bernadins**

19/05/2017 — **COLLOQUE** « La justice prédictive »
**> Cour d'Appel de Douai, Ordre des Avocats de Lille et Faculté de Droit de l'Université Catholique de Lille**

02/06/2017 — **ATELIERS** « Loyauté des traitements et décision algorithmiques »
**> LabEx Centre International de Mathématiques et Informatique de Toulouse**

08/06/2017 **DÉBAT** « Algorithmes en santé : quelle éthique ? »
**> Groupe VYV (MGEN – ISTYA – Harmonie)**

14/06/2017 **TABLE-RONDE** « Intelligence artificielle : l'éthique, à la croisée des RH et du Big Data »
**> Confédération française de l'encadrement – Confédération générale des cadres (CFE-CGC)**

16/06/2017 **DÉBAT** « Algorithmes, emploi et éthique »
**> Association française de droit du travail et de la sécurité sociale (AFDT)**

19/06/2017 **JOURNÉE** « Les algorithmes éthiques, une exigence morale et un avantage concurrentiel »
**> CERNA d'Allistene et Société Informatique de France (SIF)**

19/06/2017 **COLLOQUE** « Humain, non-humain à l'ère de l'intelligence artificielle »
**> Université Paris II**

21/06/2017 **COLLOQUE** « Intelligence artificielle : autonomie, délégation et responsabilité »
**> Ecole Nationale Supérieure de Cognitique (ENSC)**

22/06/2017 **ATELIER** « Ethique des algorithmes : enjeux pour la santé »
**> Genotoul (plateforme éthique et bioscience)**

22/06/2017 **ATELIER DE CROWDSOURCING** « Intelligence artificielle et droit »
**> OpenLaw**

22/06/2017 **COLLOQUE** « The many dimensions of data »
23/06/2017 **> Institut Mines-Télécom, Chaire de recherche Valeurs et Politiques des Informations Personnelles**

27/06/2017 **COLLOQUE** « Sécurité et justice, le défi de l'algorithme »
**> Institut national des hautes études de la Sécurité et de la Justice (INHESJ)**

28/06/2017 **PROCÈS FICTIF ET TABLE-RONDE** « Ethique, algorithmes et justice »
**> Faculté de Droit de l'Université Catholique de Lyon et Cour administrative d'appel de Lyon**

28/06/2017 **JOURNÉE D'ÉTUDES** « Admission Post-bac, cas d'école des algorithmes publics »
**> Fondation Internet Nouvelle Génération (FING) et Etalab**

03/07/2017 **JOURNÉE** « Algorithmes et souveraineté numérique »
**> CERNA d'Allistene**

05/07/2017 **JOURNÉE** « Ethique et intelligence artificielle »
**> Comité d'éthique du CNRS (COMETS) et Association française pour l'IA (AFIA)**

22/08/2017 **DÉBATS** sur les algorithmes dans le champ de l'éducation.
24/08/2017 **> Ligue de l'Enseignement**

05/09/2017 **MATINÉE-DÉBAT** « Le travail à l'ère des algorithmes : quelle éthique pour l'emploi ? »
**> Renaissance Numérique et Randstad**

11/09/2017 **COLLOQUE** « Convergences du droit et du numérique »
13/09/2017 **> Université de Bordeaux**

14/09/2017 **JOURNÉE** « Algorithmes et Politiques. Les enjeux éthiques des formes de calcul numérique vus par les sciences sociales »
**> Ecole des Hautes Etudes en Sciences Sociales (EHESS) et Institut des Systèmes Complexes Ile-de-France**

15/09/2017 **JOURNÉE** sur la recherche en santé dans ses aspects éthiques et réglementaires (données, algorithmes)
**> Hôpital Necker et Institut Imagine**

15/09/2017 **TABLES-RONDES** « Algorithmes et risques de discriminations dans le secteur de l'assurance »
**> Ligue des Droits de l'Homme**

20/09/2017 **COLLOQUE** « Enjeux éthiques des algorithmes »
**> INNOvation Ouverte par Ordinateur (INNOOO)**

20/09/2017 **MATINÉE-DÉBAT** « L'éthique des algorithmes et de l'IA est-elle compatible avec la création de valeur dans l'IoT ? : Internet of Things et/ou Internet of Trust ? »
**> ESCP Europe (Chaire IoT)**

20/09/2017 **COLLOQUE** « Ethique et numérique »
**> Collège des Bernardins**

21/09/2017 **DÉBAT** « Opportunities and challenges of advanced machine learning algorithms »
**> The John F. Kennedy Jr. Forum at Harvard Kennedy School**

21/09/2017 **COLLOQUE** « Lex Robotica (à la frontière de la robotique et du Droit : penser l'humanoïde de 2017) »
**> Conservatoire National des Arts et Métiers (CNAM)**

22/09/2017 **TABLE-RONDE** « IA et éthique des algorithmes »
**> Fotonower**

26/09/2017 **COLLOQUE** « Algorithmes prédictifs : quels enjeux éthiques et juridiques ? »
**> Centre de recherche de l'école des officiers de la gendarmerie nationale (CREOGN)**

28/09/2017 **CONSULTATION** « Quel avenir pour la médecine à l'heure de l'intelligence artificielle ? »
**> Conseil départemental du Rhône de l'Ordre des Médecins**

29/09/2017 **TABLES-RONDES** « Ethique des algorithmes et du big data »
**> Agence française de développement (AFD) et Caisse des dépôts et consignations (CDC)**

04/10/2017 **COLLOQUE** « Algorithmes et champ de bataille »
Forum-débat « Vers une Intelligence Artificielle bienveillante ? »
**> Institut des hautes études de défense nationale (IHEDN)**

06/10/2017 **FORUM-DÉBAT** « Vers une Intelligence Artificielle bienveillante ? »
**> Laboratoire d'intelligence collective et artificielle (LICA)**

12/10/2017 **TABLE-RONDE** « Droit et intelligence artificielle : quelle(s) responsabilité(s) ? »
**> Club des Juristes et Microsoft**

14/10/2017 **CONCERTATION CITOYENNE** sur les enjeux éthiques des algorithmes
**> CNIL**

07/09/2017
31/03/2018 **CONSULTATION PUBLIQUE** sur la gouvernance de l'intelligence artificielle
**> The Future Society at Harvard Kennedy School**

# GLOSSAIRE

**Algorithme**
Description d'une suite finie et non ambigüe d'étapes ou d'instructions permettant d'obtenir un résultat à partir d'éléments fournis en entrée.

**Apprentissage machine (ou apprentissage automatique, *machine learning*)**
Branche de l'intelligence artificielle, fondée sur des méthodes d'apprentissage et d'acquisition automatique de nouvelles connaissances par les ordinateurs, qui permet de les faire agir sans qu'ils aient à être explicitement programmés.

**Apprentissage machine supervisé**
L'algorithme apprend de données d'entrée qualifiées par l'humain et définit ainsi des règles à partir d'exemples qui sont autant de cas validés.

**Apprentissage machine non supervisé**
L'algorithme apprend à partir de données brutes et élabore sa propre classification qui est libre d'évoluer vers n'importe quel état final lorsqu'un motif ou un élément lui est présenté. Pratique qui nécessite que des instructeurs apprennent à la machine comment apprendre.

**Big data**
Désigne la conjonction entre, d'une part, d'immenses volumes de données devenus difficilement traitables à l'heure du numérique et, d'autre part, les nouvelles techniques permettant de traiter ces données, voire d'en tirer par le repérage de corrélations des informations inattendues.

**Chatbot**
Agent conversationnel qui dialogue avec son utilisateur (par exemple, les robots empathiques à disposition de malades, ou les services de conversation automatisés dans la relation au client).

**Intelligence artificielle (IA)**
Théories et techniques « consistant à faire faire à des machines ce que l'homme ferait moyennant une certaine intelligence » (Marvin Minsky). On distingue IA faible (IA capable de simuler l'intelligence humaine pour une tâche bien déterminée) et IA forte (IA générique et autonome qui pourrait appliquer ses capacités à n'importe quel problème, répliquant en cela une caractéristique forte de l'intelligence humaine, soit une forme de « conscience » de la machine).

**Conception & réalisation graphique :** LINÉAL - 03 20 41 40 76 / www.lineal.fr

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

Commission Nationale de l'Informatique et des Libertés

3 place de Fontenoy
TSA 80715
75334 PARIS CEDEX 07

Tél. 01 53 73 22 22
Fax 01 53 73 22 00

www.cnil.fr

et éthique numérique
les algorithmes en débat