Union internationale des télécommunications

INITIATIVE MONDIALE EN FAVEUR DE L'INCLUSION FINANCIÈRE (FIGI)

(10/2019)

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS DE
L'UIT

Groupe de travail sur la sécurité, l'infrastructure et la confiance

Mégadonnées, apprentissage automatique, protection des usagers et confidentialité

Rapport de Trust Workstream

DÉCHARGE DE RESPONSABILITÉ

L'Initiative mondiale en faveur de l'inclusion financière (FIGI) est un programme triennal mis en œuvre dans le cadre d'un partenariat entre le Groupe de la Banque mondiale, le Comité sur les paiements et les infrastructures de marché (CPMI) et l'Union internationale des télécommunications (UIT), et financé par la Bill and Melinda Gates Foundation. Il vise à faciliter et à accélérer l'application de réformes nationales en vue d'atteindre les objectifs nationaux en matière d'inclusion financière et, à terme, l'objectif mondial consistant à garantir un accès universel aux services financiers à l'horizon 2020. La FIGI finance des initiatives dans trois pays – la Chine, l'Égypte et le Mexique – et lutte contre trois obstacles distincts à l'accès financier universel, à travers le soutien qu'elle apporte aux trois groupes de travail suivants:
1) le Groupe de travail sur l'acceptation des paiements électroniques (dirigé par le Groupe de la Banque mondiale);
2) le Groupe de travail sur l'identité numérique pour les services financiers (dirigé par le Groupe de la Banque mondiale); et
3) le Groupe de travail sur la sécurité, l'infrastructure et la confiance (dirigé par l'UIT).

La FIGI organise trois colloques annuels rassemblant les autorités nationales, le secteur privé et d'autres parties prenantes compétentes afin de partager les nouvelles idées des groupes de travail et de faire le point sur l'avancée de la mise en œuvre au niveau national.

Le présent rapport a été élaboré par le Groupe de travail de la FIGI sur la sécurité, l'infrastructure et la confiance, dirigé par l'UIT. Les résultats, interprétations et conclusions exprimés dans ce rapport ne reflètent pas nécessairement les opinions des partenaires de la FIGI, notamment le CPMI, la Bill and Melinda Gates Foundation, l'UIT ou la Banque mondiale (y compris son Conseil d'administration ou les gouvernements qu'il représente). Les références éventuelles à certaines sociétés ou aux produits de certains fabricants ne signifient pas que l'UIT approuve ou recommande ces sociétés ou ces produits de préférence à d'autres de nature similaire, mais dont il n'est pas fait mention. Sauf erreur ou omission, les noms des produits propriétaires comprennent une lettre majuscule initiale. Les partenaires de la FIGI ne garantissent pas l'exactitude des données figurant dans le présent rapport. Les frontières, couleurs, dénominations et autres informations figurant sur les cartes de ce document n'impliquent aucune prise de position de la part des partenaires de la FIGI concernant le statut juridique d'un pays, d'un territoire, d'une ville ou d'une région ou de ses autorités, ni aucune reconnaissance ou acceptation de ces frontières.

Mégadonnées, apprentissage automatique, protection des usagers et confidentialité

Trust Workstream

Remerciements

Le présent rapport a été rédigé par Rory Macmillan, associé chez Macmillan Keck Attorneys & Solicitors. Il a bénéficié de la contribution et des observations de Fredesvinda Montes, de la Banque mondiale, d'Isabelle Barres, du Centre pour l'inclusion financière, de Juliet Maina, de la GSM Association (GSMA), d'Andrew Reiskind, de Mastercard, ainsi que de nombreuses autres personnes issues d'organismes fournisseurs de services financiers et d'autres organisations.

Table des matières

1 Résumé

Le présent document étudie les difficultés rencontrées dans les domaines de la législation et de la réglementation en matière de protection des usagers et de confidentialité des données en ce qui concerne les techniques liées aux mégadonnées et à l'apprentissage automatique, en particulier lorsque celles-ci sont utilisées prendre des décisions relatives aux services proposés aux usagers.

L'avantage que représentent les données pour le développement est largement reconnu, notamment pour la prestation de services financiers numériques. Les fournisseurs de services peuvent s'appuyer sur les mégadonnées pour dresser le profil personnel détaillé d'un individu, qui renseignera notamment sur son comportement (par exemple, ses préférences, ses activités et ses déplacements), lequel peut être utilisé pour adapter les offres commerciales. Les mégadonnées et l'apprentissage automatique contribuent de plus en plus à l'inclusion financière, non seulement dans les nations riches, mais aussi dans les pays en développement. Ces nouvelles technologies comportent également des risques, ou des "tendances" comme diront certains (par exemple, partialité du processus décisionnel, discrimination, atteinte à la vie privée).

L'intelligence artificielle implique des techniques qui cherchent à imiter certains aspects de la cognition humaine ou animale au moyen d'appareils informatiques. L'apprentissage automatique désigne la capacité d'un système à améliorer ses performances, en dégageant des motifs à partir de grands ensembles de données. Les mégadonnées reposent sur un traitement informatique impliquant des volumes élevés et plusieurs variétés de types de données interconnectées et traitées à grande vitesse (les "trois V", qui sont parfois au nombre de quatre avec l'ajout de l'aspect lié à la "véracité"). En règle générale, c'est également ce qui les définit.

La protection des usagers implique l'intervention de l'État, qui met en place des lois et des procédures régissant ce qui serait autrement une relation privée entre l'usager et le fournisseur. Elle vise à compenser les asymétries d'information, de négociation et de ressources perçues entre les fournisseurs et les usagers.

De plus en plus, les pays légifèrent pour protéger les données personnelles et la vie privée des personnes concernées, en leur accordant des droits qui leur donnent plus de pouvoir quant à leur utilisation. Ces lois sont mises à rude épreuve à l'ère des mégadonnées et de l'apprentissage automatique. Il est difficile de se conformer à l'obligation d'informer les usagers de la finalité de la collecte de données lorsque, comme dans le cas de l'apprentissage automatique, celle-ci n'est pas toujours connue à l'étape de notification. Le consentement est quant à lui difficile à obtenir lorsque la complexité des systèmes de mégadonnées et d'apprentissage automatique dépasse l'entendement des usagers. La notion de minimisation des données (recueillir uniquement les données nécessaires à un objectif donné, et les conserver le moins longtemps possible) va à l'encontre du modus operandi du secteur, qui met l'accent sur la maximisation des volumes de collecte de données dans le temps. Comme l'indique un rapport remis au Président des États-Unis en 2014, l'information et le consentement sont compromis par les avantages mêmes offerts par les mégadonnées: des méthodes d'utilisation nouvelles, imprévues et étonnamment puissantes des données.

Certains suggèrent que les attentes en matière de confidentialité sont fortement liées au contexte. Des restrictions plus sévères en matière de collecte, d'utilisation et de partage des données à caractère personnel dans certaines situations (et un consentement à plusieurs niveaux qui différencie les types de données en fonction de leur utilisation ou de l'entité susceptible de les utiliser) ont été envisagées. Des dispositions de temporisation prévoyant l'expiration du consentement de l'individu à l'utilisation de ses données personnelles (et un renouvellement potentiel) ont également été suggérées. Des efforts sont également déployés pour mettre au point des technologies et des services permettant de mieux gérer le

consentement. Il semble exister une véritable opportunité commerciale d'investissement et d'innovation dans l'amélioration de la gestion du consentement des usagers.

Le bon fonctionnement des modèles d'apprentissage automatique et la précision de leurs résultats dépendent de la qualité des données d'entrée. Les lois sur la protection des données et de la vie privée imposent de plus en plus aux entreprises l'obligation légale de garantir l'exactitude des données qu'elles détiennent et manipulent. Cependant, elles n'abordent pas la précision des résultats des systèmes de mégadonnées et d'apprentissage automatique. Cela soulève plusieurs questions quant aux responsabilités réglementaires de ceux qui manipulent les mégadonnées, concernant l'exactitude i) des données d'entrée dans les décisions automatisées ainsi que ii) des données rapportées dans les systèmes officiels de communication des données de crédit. Dans certaines juridictions, cette incertitude a donné lieu, entre autres recours, à certains droits d'opposition aux décisions automatisées.

Les déductions tirées des données d'entrée générées par les modèles d'apprentissage automatique déterminent la manière dont les individus sont perçus et évalués dans le cadre de décisions automatisées. Les lois sur la protection des données et de la vie privée peuvent s'avérer insuffisantes pour couvrir les résultats des modèles d'apprentissage automatique qui traitent ces données. L'une de leurs priorités est de prévenir la discrimination, en protégeant notamment certaines catégories de groupes (en fonction, par exemple, de l'appartenance ou de l'origine ethnique, de la religion ou du genre des personnes). Cependant, à l'ère des mégadonnées, des données non sensibles peuvent être utilisées pour déduire des données sensibles.

L'apprentissage automatique peut aboutir à des résultats discriminatoires lorsque l'apprentissage des algorithmes repose sur des exemples historiques qui reflètent des discriminations passées, ou que le modèle ne prend pas en compte un ensemble de facteurs suffisamment large. Il s'avère difficile de surmonter les biais, mais des tests ont été mis au point afin d'évaluer où ceux-ci peuvent survenir. Dans certains pays, un biais, même non intentionnel, peut être illégal s'il a un "impact disparate", c'est-à-dire lorsque les résultats d'un processus de sélection sont très différents pour une catégorie de personnes protégées.

La question est de savoir dans quelle mesure les entreprises doivent assumer la responsabilité et le coût de l'identification des discriminations et des biais potentiels dans leurs algorithmes de données. Les entreprises qui s'appuient sur les mégadonnées et l'apprentissage automatique pourraient employer des outils pour s'assurer que leurs données n'exacerberont pas des préjugés historiques et utiliser celles-ci pour repérer les discriminations (dans certaines juridictions, cela relève d'ailleurs de leur responsabilité). Des cadres déontologiques et l'adoption de "bonnes pratiques" peuvent être nécessaires pour garantir le suivi et l'évaluation des résultats, ainsi que l'ajustement des algorithmes.

Les grands volumes de données détenues et transférées par les acteurs du domaine des mégadonnées risquent de compromettre la sécurité des données, et donc d'entraîner des risques pour la vie privée des usagers. La vie privée peut être protégée à des degrés divers grâce à l'utilisation de technologies d'amélioration de la confidentialité. Le marché des services de désidentification, de pseudonymisation et d'anonymisation est en pleine expansion. La confidentialité différentielle est également de plus en plus utilisée. Il pourrait s'avérer nécessaire de mettre en place une réglementation afin de veiller à ce que les technologies d'amélioration de la confidentialité soient systématiquement intégrées dans le traitement des mégadonnées et des données de l'apprentissage automatique. Cette démarche peut nécessiter l'établissement de mesures incitatives dans le cadre de la législation qui impliquent une responsabilité en cas de violation des données, faisant essentiellement peser le fardeau économique non pas sur les usagers

en obtenant leur consentement, mais sur les organisations qui recueillent, utilisent et partagent les données.

Les mégadonnées et l'apprentissage automatique sont rendus possibles par des intermédiaires, tels que les courtiers en données indépendants qui font le commerce des données personnelles. Le transfert de données à caractère personnel entraîne plusieurs risques: violation de données et usurpation d'identité, marketing intrusif et autres atteintes à la vie privée. Tandis que les courtiers en données font l'objet d'un examen de plus en plus minutieux, des lois qui accordent des droits directs aux usagers sont en cours d'adoption.

Les exigences traditionnelles consistant à renseigner l'usager sur la finalité de l'utilisation des données à caractère personnel alors que celle-ci est encore incertaine, ou à obtenir son consentement à l'égard de quelque chose qui dépasse de loin son entendement, sont remises en question. Les risques liés à l'inexactitude des données saisies ou au traitement partial et discriminatoire dans les processus décisionnels automatisés soulèvent également plusieurs questions complexes, par exemple, sur la manière de garantir que les consommateurs ne sont pas traités injustement. Assurer la transparence des décisions générées par les algorithmes ou mettre en évidence les préjudices qui ont été directement provoqués par les technologies de l'intelligence artificielle représente également un défi dans le cadre de la législation et de la réglementation en matière de protection des usagers et de confidentialité des données.

Les défis posés par le traitement des mégadonnées et des données de l'apprentissage automatique dans les cadres juridiques et réglementaires de la protection des données et de la vie privée portent à croire que la mise en place de solides systèmes d'autorégulation et d'éthique au sein de la communauté de l'intelligence artificielle et des services financiers pourrait s'avérer particulièrement importante. Face à l'incertitude juridique et réglementaire, les entreprises peuvent mettre en place des systèmes de gestion des risques, appliquer le principe de confidentialité dès la conception et élaborer un cadre déontologique.

Il existe plusieurs domaines qui méritent d'être approfondis et de faire l'objet de normes et de procédures, notamment en ce qui concerne les analyses déductives acceptables, la fiabilité des déductions, les normes éthiques relatives à l'intelligence artificielle, la communication d'éléments contrefactuels en aval de la prise de décisions, la documentation des politiques écrites, l'intégration des principes liés à la confidentialité à la conception, les explications justifiant les décisions automatisées, l'accès à l'intervention humaine et d'autres mécanismes de redevabilité.

2 Acronymes

AIDA	Intelligence artificielle et analyse des données
APEC	Coopération économique Asie-Pacifique
API	Interface de programmation d'applications
CGAP	Groupe consultatif pour l'aide aux plus pauvres
FCRA	Fair Credit Reporting Act
FEAT	Équité, éthique, responsabilité et transparence
GPFI	Partenariat mondial pour l'inclusion financière
IEEE	Institut d'ingénierie électrique et électronique
OCDE	Organisation de coopération et de développement économiques
RGPD	Règlement général sur la protection des données de l'Union européenne

3 Introduction

Les mégadonnées, l'intelligence artificielle et l'apprentissage automatique dominent le débat public, qu'il s'agisse de l'enthousiasme suscité par les nouvelles fonctionnalités ou des craintes liées aux pertes d'emplois et aux préjugés découlant d'un processus décisionnel automatisé. Ces questions ne sont pas tout à fait nouvelles[1]. Cependant, la sensibilisation du public au potentiel des systèmes informatiques puissants qui utilisent des algorithmes complexes pour traiter d'énormes volumes de données s'est accrue, tandis que les ordinateurs battent désormais les êtres humains à divers jeux et que les individus apprécient de plus en plus les services assurés par ces systèmes[2].

Les données personnelles identifiables sont recueillies à grande échelle, partagées et disponibles sur les marchés commerciaux de données. Ces données peuvent concerner l'historique des transactions et de navigation sur Internet d'une personne, son inscription auprès d'organisations publiques et privées, et son utilisation des médias sociaux. Les entreprises et les gouvernements recueillent, traitent et partagent régulièrement ces données avec des parties tierces, souvent à l'insu de la personne concernée et sans son consentement.

L'avantage que représentent les données pour le développement est largement reconnu, notamment pour la prestation de services financiers numériques[3]. De nombreux services financiers dépendent de l'évaluation et de la gestion des risques. Par exemple, la valeur d'un prêt repose en grande partie sur la solvabilité de l'emprunteur, ainsi que sur les garanties relatives au prêt. Plus il y a de données sur l'emprunteur, mieux le prêteur peut évaluer sa solvabilité. Les mégadonnées permettent de tirer des conclusions sur la solvabilité d'un emprunteur sur la base de son appartenance à une ou plusieurs catégories de personnes qui ont emprunté et remboursé, ou failli à le faire, leurs dettes par le passé.

Les fournisseurs de services financiers numériques peuvent non seulement générer des bénéfices commerciaux, mais aussi, grâce aux informations et à l'analyse des antécédents et des intérêts des usagers, apporter une valeur ajoutée publique substantielle en améliorant l'accès à ces services.

L'intelligence artificielle est de plus en plus utilisée dans l'analyse d'un large éventail de sources de données afin d'évaluer avec cohérence la solvabilité des consommateurs et de prendre des décisions relatives aux prêts. Plutôt que de se contenter de la représentation des revenus et des dettes de l'emprunteur dans la demande de prêt ou d'un entretien avec le directeur de la banque locale, ou de vérifier le score accordé par une agence d'évaluation du crédit (par exemple, FICO), les entreprises sont en mesure d'analyser l'empreinte numérique d'un individu en vue de prédire la probabilité de défaut de paiement grâce à l'intelligence artificielle et aux mégadonnées. Cela permet d'accéder à des services qui, autrement, n'auraient peut-être pas été disponibles.

L'analyse des mégadonnées peut contribuer à améliorer les moyens traditionnels d'évaluation du crédit. Les agences d'évaluation du crédit, telles qu'Equifax, ont affirmé avoir considérablement amélioré la

---

[1] Cela fait plus de 20 ans que le risque de biais dans les systèmes informatiques, et les solutions pour y remédier, font l'objet de discussions. Voir: Friedman, B. et Nissenbaum, H., "Bias in Computer Systems". *ACM Transactions on Informations Systems*, vol. 14, nº 330, 1996.

[2] Voir, par exemple: Metz, C., "In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go". WIRED, 27 janvier 2016. Disponible à l'adresse suivante: https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/.

[3] Groupe consultatif d'experts indépendants du Secrétaire général des Nations Unies sur la révolution des données pour le développement durable, *A World That Counts: Mobilizing the Data Revolution for Sustainable Development*.

capacité prédictive de leurs modèles en recourant à l'analyse des mégadonnées. Cela peut être particulièrement utile pour évaluer les individus sans antécédents traditionnels en matière de crédit, qui peuvent ainsi avoir accès aux services de crédit. Outre l'amélioration des moyens traditionnels d'évaluation du crédit, l'analyse des mégadonnées s'étend à des modèles entièrement nouveaux. La plate-forme Upstart[4], par exemple, utilise l'apprentissage automatique pour prédire la solvabilité des jeunes adultes à partir de données relatives à leur éducation, leurs résultats aux examens, leur domaine d'études et leurs antécédents professionnels, dans le cadre d'un processus de prêt automatisé. Elle propose des prêts directement aux consommateurs et met son logiciel à la disposition d'autres prêteurs, qui bénéficient ainsi d'une plate-forme pour leurs propres services de prêt.

Ces modèles commerciaux contribuent de plus en plus à l'inclusion financière, non seulement dans les nations riches, mais aussi dans les pays en développement. Lenndo[5], une fintech appuyant l'évaluation du crédit avec une analyse alternative des données, s'est associée à FICO, l'agence mondiale d'évaluation du crédit, pour rendre ses services de notation disponibles en Inde[6]. Ceux-ci permettent d'évaluer, en collaboration avec l'une des agences de crédit indiennes, les données alternatives de l'empreinte numérique des consommateurs afin d'attribuer un score de crédit à ceux qui n'ont pas suffisamment de données traditionnelles dans leur dossier ("dossier léger") en vue d'une approbation de prêt traditionnelle. Actives en Afrique et ailleurs, les entreprises Branch.co[7] et MyBucks[8] utilisent le contrôle d'identité et une application mobile automatisée qui se sert des moteurs d'évaluation du crédit pour générer des scores de crédit à partir de l'analyse de la facture de téléphone mobile, de messages textes, de l'historique des paiements ou des comptes bancaires (le cas échéant), des factures de services publics ainsi que des données de géolocalisation des consommateurs.

L'accès rapide à de grands volumes de données est la clé de l'efficacité de ces technologies. Par exemple, la société ZestFinance a conclu un accord stratégique avec son investisseur Baidu, le fournisseur chinois de services de recherche sur Internet (l'équivalent de Google en Chine), qui lui permet d'accéder à l'historique de navigation, ainsi qu'aux données de géolocalisation et de paiement des particuliers pour établir des scores de crédit en Chine, où près de la moitié de la population n'a pas d'antécédents de crédit[9]. Pour reprendre les célèbres propos du directeur général de ZestFinance, toutes les données sont des données de crédit[10].

L'intelligence artificielle n'est pas seulement utile pour l'évaluation des risques en matière de crédit. Tout service impliquant une évaluation des risques repose sur les informations et l'analyse. La société Progressive, par exemple, recueille des données sur les performances individuelles des conducteurs au moyen d'applications mobiles telles que Snapshot, afin de prédire les risques d'accident et de proposer (le cas échéant) des primes d'assurance à prix réduit[11]. De nombreuses autres applications

[4] https://www.upstart.com/.
[5] https://www.lenddo.com/.
[6] http://www.prnewswire.co.in/news-releases/new-fico-credit-scores-provide-lenders-opportunity-to-expand-access-to-credit-in-india-for-nearly-350-million-653029163.html.
[7] https://branch.co/.
[8] https://www.mybucksbanking.mw/
[9] https://www.businesswire.com/news/home/20160717005040/en/ZestFinance-Receives-Funding-Baidu-Fuel-Development-Search-Based.
[10] https://www.pymnts.com/in-depth/2015/how-zestfinance-used-big-data-lending-to-secure-150m-from-fortress/.
[11] https://www.progressive.com/auto/discounts/snapshot/.

du domaine des assurances ont recours à l'intelligence artificielle[12]. Parmi les autres domaines où elle influe grandement sur l'innovation et l'efficacité, citons la personnalisation des produits d'épargne, la gestion des services de paiement, l'assistance virtuelle aux consommateurs (par exemple, les robots conseillers et les agents conversationnels), ainsi que la détection de la fraude, du blanchiment d'argent et du financement du terrorisme.

L'augmentation de l'utilisation par les consommateurs de produits et de services reposant sur l'intelligence artificielle et l'apprentissage automatique a déclenché un débat politique véhément sur les risques connexes et la nécessité de mettre en œuvre des mesures politiques cohérentes[13]. La Banque mondiale a préparé un rapport pour le Sommet du G20 de 2018, intitulé *Use of Alternative Data to Enhance Credit Reporting to Enable Access to Digital Financial Services by Individuals and SMEs operating in the Informal Economy* (Utilisation de données alternatives pour améliorer l'évaluation du crédit afin de permettre l'accès aux services financiers numériques des particuliers et des petites et moyennes entreprises opérant dans l'économie informelle)[14]. Le présent rapport s'appuie sur les grandes questions analysées dans le document susmentionné, ainsi que sur les recommandations formulées dans le cadre de celui-ci à l'intention des décideurs et des organismes de régulation.

L'Autorité monétaire de Singapour a récemment publié des principes visant à promouvoir l'équité, l'éthique, la responsabilité et la transparence dans l'utilisation de l'intelligence artificielle et l'analyse des données dans le secteur financier du pays (*Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*). L'objectif est d'appliquer les principes liés à l'équité, à la responsabilité et à la transparence spécifiquement au contexte de l'intelligence artificielle et de l'apprentissage automatique dans le secteur financier, en ajoutant une dimension éthique. Les principes FEAT sont présentés à l'annexe A (Principes FEAT de l'Autorité monétaire de Singapour). La Smart Campaign a récemment publié la première ébauche du document intitulé *Digital Credit Standards*, qui porte sur les normes relatives au crédit numérique et comprend un certain nombre de normes relatives à l'utilisation des données, au profilage et aux décisions automatisées dans les services financiers numériques, lesquelles sont présentées à l'annexe B (Normes de la Smart Campaign relatives au crédit numérique). Il y sera fait référence de temps à autre dans le présent rapport pour illustrer la manière dont les questions concernant la protection des usagers peuvent être abordées.

L'Initiative mondiale de l'Institut d'ingénierie électrique et électronique (IEEE) pour les considérations éthiques dans l'intelligence artificielle et les systèmes autonomes recommande ce qui suit[15]:

> *Les législateurs nationaux, et surtout internationaux, doivent être encouragés à envisager et à étudier soigneusement la nécessité éventuelle d'introduire une nouvelle réglementation, le cas échéant, y compris des règles qui soumettent le lancement sur le marché de nouvelles technologies*

---

[12] Voir la page suivante : https://www.techemergence.com/machine-learning-at-insurance-companies/.
[13] Voir, par exemple: Calo, R., "Artificial Intelligence Policy: A Primer and Roadmap". Disponible à l'adresse suivante: https://ssrn.com/abstract=3015350.
[14] Partenariat mondial pour l'inclusion financière (GPFI), "Use of Alternative Data to Enhance Credit Reporting to Enable Access to Digital Financial Services by Individuals and SMEs operating in the Informal Economy", note d'orientation élaborée par l'International Committee on Credit Reporting. (Document exposant les priorités du GPFI de 2018). Disponible à l'adresse suivante: https://www.g20.org/sites/default/files/documentos_producidos/use_of_alternative_data_to_enhance_credit_reporting_to_enable_access_to_digital_financial_services_iccr.pdf.
[15] Voir *Ethically Aligned Design*, à la note de bas de page 223.

reposant sur l'intelligence artificielle en tant que service à l'examen et à l'approbation préalables des organismes nationaux ou internationaux appropriés.

Les lois et réglementations de longue date qui visent à protéger les usagers contre les utilisations préjudiciables de leurs données personnelles sont mises à mal par de nouvelles méthodes de collecte et d'analyse des données. En effet, certains se sont risqués à dire que même la plus récente des lois sur la protection des données et de la vie privée, le Règlement général sur la protection des données (RGPD) de l'Union européenne, parfois qualifiée de "référence" en la matière, est "incompatible" avec le monde des mégadonnées[16]. Des préoccupations similaires se posent en ce qui concerne d'autres normes mondiales, notamment la Recommandation de l'Organisation de coopération et de développement économiques (OCDE) concernant les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel (les Lignes directrices de l'OCDE sur la protection de la vie privée)[17] et la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (Série des traités européens nº 108, dénommée Convention 108), telle que récemment modifiée par le Protocole d'amendement à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel[18].

Les trois principes fondamentaux de la loi sur la protection des données et de la vie privée sont la spécification de la finalité, la minimisation des données et le traitement des données des catégories de groupes dites "protégées" ou "particulières" (personnes d'une certaine appartenance ou origine ethnique, d'une certaine religion, d'un certain genre, etc.). Ces principes sont mis à rude épreuve lorsque l'objectif spécifique de la collecte et du traitement des données ne peut être compris qu'au fur et à mesure que les ordinateurs eux-mêmes apprennent à partir de gros volumes de données d'observation et de performance, produisant ainsi une analyse plus précise. Les données personnelles peuvent également indiquer l'appartenance à un groupe protégé.

Ces nouvelles technologies comportent également des risques, ou des "tendances" comme diront certains (par exemple, partialité du processus décisionnel, discrimination, atteinte à la vie privée)[19]. L'analyse des données peut être utilisée pour tirer des conclusions (et dans certains cas faire des prédictions) sur les attributs d'une personne: son appartenance ethnique, son genre, son orientation sexuelle, ses relations, ses opinions politiques, sa santé (y compris certaines maladies), son état mental, ses intérêts personnels, sa solvabilité, etc. La discrimination peut alors faire partie intégrante du traitement des données,

---

[16] Les dispositions du RGPD sont – pour reprendre un terme clé utilisé dans l'ensemble du document – incompatibles avec l'environnement informatique créé par la disponibilité des mégadonnées. Une telle incompatibilité ne peut qu'invalider rapidement de nombreuses dispositions du RGPD. Zarsky, T., "Incompatible: The GDPR in the Age of Big Data". *Seton Hall Law Review*, vol. 47, nº 4 (2), 8 août 2017. Disponible à l'adresse suivante: https://ssrn.com/abstract=3022646.

[17] Voir la page suivante : https://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf.

[18] Le Protocole d'amendement à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, qui a modifié la Convention 108 en 2018, aborde désormais différentes fonctionnalités du traitement automatisé des données, telles que le profilage, les décisions automatisées et l'utilisation d'algorithmes. Il s'agit notamment du droit de ne pas faire l'objet d'une décision ayant une incidence significative, fondée uniquement sur un traitement automatisé des données personnelles sans tenir compte du point de vue de la personne concernée; du droit d'avoir connaissance de la logique qui sous-tend le traitement des données lorsque les résultats sont appliqués à la personne concernée; et du droit de s'opposer au traitement de ses données personnelles.

[19] Citron, D. et Pasquale, F. A., "The Scored Society: Due Process for Automated Predictions". Publication du Social Science Research Network (SSRN) nº 2376209, 2014. Disponible à l'adresse suivante: https://papers.ssrn.com/abstract=2376209; Zarsky, T. Z., "Understanding Discrimination in the Scored Society". *Washington Law Review*, vol 89, nº 1375, 2014; Mittelstadt, B., *et al.*, "The Ethics of Algorithms: Mapping the Debate". *Big Data & Society*, vol. 3, 2016. Disponible à l'adresse suivante: http://bds.sagepub.com/lookup/doi/10.1177/2053951716679679.

conduisant effectivement à des résultats qui seraient interdits par les lois sur la discrimination raciale ou basée sur le genre si les décisions étaient prises par des humains (et non par des ordinateurs).

Ces risques sont particulièrement importants pour les services financiers. Contrairement à de nombreux produits et services de consommation, les offres et les prix des services financiers dépendent du profil de chaque usager. Les décisions d'offrir un prêt (et à quel taux d'intérêt), d'émettre une carte de crédit (et avec quelle limite de crédit) et de proposer différents types d'assurance dépendent toutes de l'évaluation du risque que présente l'individu. Ainsi, tout comme la décision d'employer quelqu'un, de nombreux services financiers ont une dimension personnelle importante[20].

Il est ainsi possible de mieux adapter les services au profil de risque de chaque personne, et donc de faciliter l'accès à des services financiers qui n'auraient peut-être pas été proposés autrement. Dans le même temps, il arrive toutefois que la personne concernée n'ait pas connaissance des données ayant servi à la prise de décisions ni de la raison pour laquelle les services lui ont été refusés, et qu'elle ne soit pas en mesure de contester lesdites données, déductions et décisions.

Si l'accès aux données à caractère personnel permet de prendre de plus en plus de décisions en fonction des comportements individuels, il existe un risque d'atteinte à la vie privée. En 2008, la Commission fédérale du commerce des États-Unis est intervenue pour mettre fin aux pratiques déloyales de l'entreprise CompuCredit, qui proposait des cartes de crédit aux personnes ayant contracté un prêt à risque. CompuCredit réduisait les limites de crédit des consommateurs sur la base d'un modèle qui faisait baisser leur score lorsqu'ils effectuaient certaines transactions (visite de prêteurs sur gages, de conseillers personnels et de salles de billard, par exemple)[21].

Le traitement des données personnelles disponibles, en particulier le profilage et le processus visant à faire des déductions au sujet d'une personne, est donc essentiel à la prestation de ces services financiers. Par conséquent, pour garantir l'équité, l'exactitude et la transparence dans le cadre des services financiers, il convient de tenir compte de la nature des données personnelles et de la manière dont elles sont recueillies, utilisées et partagées avec des parties tierces[22].

Ces difficultés sont exacerbées par la variété des cadres réglementaires s'appliquant aux différents types de fournisseurs de services financiers numériques (certains sont réglementés comme des banques, tandis que d'autres le sont à peine). Différentes restrictions peuvent régir la collecte et l'utilisation des données dans le cadre de services similaires, et les consommateurs peuvent disposer de recours différents.

---

[20] Ce n'est pas le cas de tous les services financiers; par exemple, il n'y a aucune raison particulière de traiter les investisseurs différemment dans le cadre des dépôts bancaires des petits épargnants ou d'un fonds d'investissement accessible aux investisseurs de détail.

[21] Singel, R., "Credit Card Firm Cut Limits After Massage Parlor Visits, Feds Allege", *Wired*, 20 juin 2008, disponible à l'adresse suivante: https://www.wired.com/2008/06/credit-card-fir/; Réclamation de la Commission fédérale du commerce des États-Unis, disponible à l'adresse suivante: https://www.wired.com/images_blogs/threatlevel/files/compucreditcmplt.pdf; "Subprime Credit Card Marketer to Provide At Least $114 Million in Consumer Redress to Settle FTC Charges of Deceptive Conduct", 19 décembre 2008, disponible à l'adresse suivante: https://www.ftc.gov/news-events/press-releases/2008/12/subprime-credit-card-marketer-provide-least-114-million-consumer; et Commission fédérale du commerce des États-Unis contre CompuCredit Corporation et Jefferson Capital Systems LLC, "Stipulated Order for Permanent Injunction and Other Equitable Relief Against Defendant CompuCredit Corporation", affaire civile nº 1:08-CV-1976-BBM-RGV, disponible à l'adresse suivante: https://www.ftc.gov/sites/default/files/documents/cases/2008/12/081219compucreditstiporder.pdf.

[22] Voir, plus généralement: Banque mondiale, *New Forms of Data Processing Beyond Credit Reporting: Consumer and Privacy Aspects*, 2018; et Responsible Finance Forum, *Opportunities and Risks in Digital Financial Services: Protecting Consumer Data and Privacy*, 2017.

Les défis posés par le traitement des mégadonnées et des données de l'apprentissage automatique dans les cadres juridiques et réglementaires de la protection des données et de la vie privée portent à croire que la mise en place de solides systèmes d'autorégulation et d'éthique au sein de la communauté de l'intelligence artificielle et des services financiers pourrait s'avérer particulièrement importante.

Le présent document fournit des informations de référence aux décideurs politiques, aux organismes de régulation, aux fournisseurs de services financiers numériques, aux investisseurs ainsi qu'à d'autres organisations au sujet des solutions et des normes nécessaires à la protection de la confidentialité des données des usagers dans le contexte des mégadonnées et de l'apprentissage automatique. Ces problèmes surviennent à mesure que les technologies, les cas d'utilisation et l'adoption des services se multiplient. Ainsi, bien que l'on comprenne de mieux en mieux les défis émergents, il existe peu de domaines dans lesquels on observe un consensus sur les meilleures pratiques. L'approche à suivre dépendra de la manière dont les décideurs politiques, les législateurs, les organismes de régulation et les acteurs du marché évalueront les compromis et les synergies entre les objectifs politiques, tels que l'expérimentation et l'innovation, la productivité économique, la confiance dans les services et la protection des usagers.

Le présent document explore différents points de vue, exposant les suggestions d'organisations, d'universitaires et de théoriciens au sujet des approches communément adoptées pour protéger la confidentialité des données des usagers, ainsi que les lois et réglementations connexes. L'objectif est de mettre ces idées en avant, et non de prendre position. Les auteurs cherchent à aider ceux qui devront s'attaquer à ces questions aux niveaux politique, législatif et réglementaire dans les années à venir. Plutôt que de recommander de bonnes pratiques à adopter, le présent document se concentre donc sur l'identification et la définition des principales questions à prendre en compte lors de l'élaboration de cadres réglementaires (y compris, éventuellement, de cadres d'autorégulation)[23].

La section 4 présente les grands concepts en jeu, en commençant par les tendances technologiques et commerciales des mégadonnées et de l'apprentissage automatique (section 4.1), les types de données utilisés (section 4.2), ainsi que le profilage et les décisions automatisées qui s'appuient sur ces données (section 4.3). Elle explique ensuite ces concepts et aborde les dimensions réglementaires mentionnées tout au long du document: la protection des usagers (section 4.4) et la confidentialité des données (section 4.5).

Le présent document s'intéresse ensuite à la protection des usagers et à la confidentialité des données au cours des trois grandes phases de l'interaction entre les usagers et les fournisseurs de services qui ont recours aux mégadonnées et à l'apprentissage automatique.

La section 5 porte sur la phase de pré-engagement, qui concerne principalement les processus visant à informer les usagers de la manière dont leurs données personnelles seront recueillies, utilisées et communiquées à des parties tierces, et à quelles fins, ainsi que sur les exigences relatives à l'obtention de leur consentement, nécessaire pour légitimer l'utilisation des données personnelles.

[23] Le présent document ne traite pas de toutes les questions relatives à la confidentialité des données et à la protection des usagers qui pourraient se poser dans le domaine des données personnelles, ni de tous les aspects des mégadonnées et de l'apprentissage automatique. Les techniques étudiées ont abouti à des résultats généraux, qui concernent l'ensemble de la société et diverses politiques (santé, éducation, etc.), mais qui n'ont pas d'incidence directe dans le cadre des décisions prises au sujet d'individus spécifiques. Certains droits et obligations sont donc étudiés de façon plus approfondie, notamment les domaines où les mégadonnées et l'apprentissage automatique posent des défis particuliers en matière de confidentialité des données et de protection des usagers.

La section 1 porte quant à elle sur la phase d'engagement, qui concerne les restrictions, les exigences et la responsabilité liées à l'utilisation des données à caractère personnel par les entreprises, notamment s'agissant de la précision des modèles d'apprentissage automatique (section 6.1), du traitement partial et discriminatoire des données (section 6.2), de la violation des données et de la réidentification (section 6.3), ainsi que de la communication des données à des parties tierces (section 1.1).

La section 7 s'intéresse à la phase de post-engagement et aux moyens dont disposent les usagers pour tenir les opérateurs de mégadonnées et des systèmes d'apprentissage automatique responsables du non-respect des lois relatives à la protection des usagers et à la confidentialité des données. Elle examine i) les droits des usagers à accéder aux données à caractère personnel les concernant, à rectifier les erreurs qu'elles contiennent et à demander leur effacement (section 7.1); ii) le manque de transparence lié à l'obtention d'explications sur les résultats complexes des modèles d'apprentissage automatique (section 7.2); iii) le droit de contester les décisions et d'obtenir une intervention humaine (section 7.3); et iv) la difficulté de démontrer un préjudice (section 7.4).

La section 8 examine certaines mesures pratiques que les entreprises pourraient prendre pour réduire les risques face aux incertitudes juridiques et réglementaires. Le présent document se clôt sur la section 0, qui dresse une courte liste d'aspects à améliorer dans ce domaine, qu'il s'agisse de l'élaboration de principes déontologiques, de normes ou de procédures.

4 Comprendre les mégadonnées, la protection des usagers et la confidentialité des données.

4.1 Que sont les mégadonnées et l'apprentissage automatique?

L'***intelligence artificielle*** implique des techniques qui cherchent à imiter certains aspects de la cognition humaine ou animale au moyen d'ordinateurs. L'***apprentissage automatique,*** une forme d'intelligence artificielle, désigne la capacité d'un système à améliorer ses performances, en dégageant des formes ou des motifs à partir de grands ensembles de données, et ce, à plusieurs

Figure 1 – Apprentissage automatique, xkcd.com

niveaux d'analyse (on parle souvent d'apprentissage profond)[24].

Les algorithmes d'apprentissage automatique bâtissent un modèle à partir de ***données d'apprentissage***, c'est-à-dire d'exemples passés, afin de faire des prédictions ou de prendre des décisions, plutôt que de suivre uniquement une logique préprogrammée. Les réseaux neuronaux analysent les données à travers de nombreuses couches de matériel et de logiciel[25]. Chaque couche produit sa propre représentation des données et partage ce qu'elle a "appris" avec la couche suivante. L'apprentissage automatique est un apprentissage par l'exemple: on utilise les données d'apprentissage pour entraîner le modèle à se comporter d'une certaine manière[26]. Cette technique n'est pas nouvelle, mais grâce aux mégadonnées, elle a trouvé de nombreuses utilisations concrètes[27].

Les **mégadonnées** reposent sur un traitement informatique impliquant des volumes élevés et plusieurs variétés de types de données interconnectées et traitées à grande vitesse (les "trois V"[28], qui sont parfois au nombre de quatre avec l'ajout de l'aspect lié à la "véracité"). En règle générale, c'est également ce qui les définit[29]. L'avènement des technologies ayant recours aux mégadonnées découle de l'évolution de la manière dont les données sont recueillies, conservées et utilisées. Les données sont recueillies à l'aide d'un grand nombre d'applications et de capteurs qui enregistrent les communications, les transactions et les déplacements des usagers. Des bases de données réparties stockent les données et des systèmes de communication à haut débit assurent leur transmission à grande vitesse, ce qui réduit le coût de l'analyse des données. Ces processus analytiques avancés sont appliqués dans de nombreux contextes.

4.2 Quel type de données est utilisé?

Par le passé, les données utilisées pour la prise de décisions dans le cadre des services financiers pouvaient inclure les déclarations formelles d'un candidat à un service, les connaissances personnelles du directeur d'une banque locale ou d'un courtier en assurance, et un éventail plus large de données organisées détenues, analysées et profilées par les agences d'évaluation du crédit. De nos jours, les mégadonnées comprennent des ***données alternatives***, c'est-à-dire des données qui ne sont pas recueillies et documentées conformément aux évaluations du crédit traditionnelles, mais qui proviennent d'un large éventail d'autres sources numériques.

Données de télécommunication

Les services des opérateurs de réseaux de télécommunication constituent une source importante de données alternatives utilisées pour étendre les services financiers. Les entreprises de télécommunication sont généralement limitées dans leur capacité à recueillir et utiliser les données relatives à leurs clients, notamment le contenu de leurs appels téléphoniques. Celles-ci ont été protégées par des lois relatives à

---

[24] Surden, H., "Machine Learning and the Law". *Washington Law Review*, vol 89, 2014, p. 87-88.
[25] Rosenblatt, F., "Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain". *Psychological Review*, vol. 65, nº 6, 1958. Disponible à l'adresse suivante: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.335.3398&rep=rep1&type=pdf.
[26] Les exemples les plus connus sont Watson (IBM), AlphaGo (Google/DeepMind), Siri (Apple) et Alexa (Amazon), qui s'appuient tous sur l'apprentissage automatique pour améliorer l'expérience utilisateur.
[27] Stone, P., *et al.*, *Artificial Intelligence and Life in 2030: Report of the 2015 Study Panel*. Stanford University, 2016. Disponible à l'adresse suivante: https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100report10032016fnl_singles.pdf .
[28] Laney, D., "3D Data Management: Controlling Data Volume, Velocity and Variety", note de recherche de META Group. 6 février 2001.
[29] IBM, "The Four V's of Big Data". 2014. Disponible à l'adresse suivante: http://www.ibmbigdatahub.com/infographic/four-vs-big-data; Mayer-Schönberger, V. et Cukier, K., *Big Data: A Revolution That Will Transform How We Live, Work and Think*, John Murray, 2013.

l'interception légale, qui portent sur des questions similaires aux lois protégeant les communications postales, qui interdisaient l'ouverture des enveloppes en l'absence de base légale. Si les entreprises de télécommunication ne peuvent pas utiliser le contenu des communications de leurs clients, elles ont toutefois accès aux métadonnées (et sont souvent tenues par la réglementation de les conserver[30]).

Les métadonnées sont des données relatives à l'utilisation des services de communication – qui a communiqué avec qui, à quelle heure, pendant combien de temps, où l'appel a été passé, etc., autant d'éléments qui peuvent permettre d'établir le profil des relations et des flux financiers d'un individu. Le rechargement régulier d'un crédit téléphonique prépayé peut laisser croire que la personne concernée dispose de revenus stables. Les appels vers et depuis l'étranger peuvent quant à eux impliquer l'accès à un réseau international, et éventuellement une certaine aisance. Au cours de la journée de travail, des appels réguliers provenant d'une zone urbaine dense suggèrent que la personne concernée bénéficie d'un emploi stable, le soir, des appels réguliers passés au même endroit peuvent indiquer l'emplacement du domicile de la personne, et donc sa catégorie économique ou sociale.

Les lois et les licences de télécommunication de nombreux pays comportent des clauses interdisant expressément aux titulaires de licence d'utiliser, de divulguer ou d'enregistrer toute communication ou tout contenu envoyé à l'aide d'un service de communication électronique, ou de transmettre toute information relative à ces services à autrui. Ce phénomène s'étend de plus en plus aux métadonnées. Par exemple, la directive européenne sur la protection de la vie privée en ligne est remplacée par le Règlement sur la protection de la vie privée en ligne, qui aborde plus en détail les questions liées à la protection des données du RGPD, notamment en ce qui concerne les services de communication électronique, en traitant à la fois des données personnelles et des métadonnées, telles que les enregistrements détaillés des appels[31]. Toutefois, cette situation n'est pas universelle, et de nombreux pays n'empêchent pas l'utilisation des métadonnées. Même lorsque celle-ci est interdite, elle peut être autorisée avec le consentement du client, ce qui permet à l'opérateur de générer des scores de crédit, qui peuvent être utilisés pour accorder des prêts numériques.

Données relatives à l'argent mobile et autres informations de paiement

Les entreprises de télécommunication peuvent également détenir des données sur l'utilisation de services connexes qui sont assurés par le biais des réseaux de télécommunication. Par exemple, de nombreux opérateurs de réseaux mobiles proposent un service de paiement mobile propriétaire à leurs clients. Ils ont ainsi accès à des données sur le moment, la régularité et le montant du rechargement du portefeuille d'argent mobile d'une personne, le solde moyen qu'elle maintient, l'origine des paiements reçus ou les bénéficiaires des transferts effectués, ainsi que le montant de ces derniers. En analysant la régularité, les montants et les bénéficiaires des paiements (par exemple, la famille), l'analyse des données peut fournir un aperçu de l'ampleur et de la fiabilité des flux de trésorerie d'une personne (ses revenus et ses dépenses), ainsi que de son réseau social et, en définitive, permettre d'évaluer sa solvabilité. Le paiement régulier

[30] Par exemple, la loi australienne de 2015 sur la conservation des données oblige les opérateurs de télécommunication et les fournisseurs d'accès à Internet à conserver les métadonnées relatives aux communications pendant deux ans, afin d'aider les forces de l'ordre dans leurs enquêtes sur la criminalité et le terrorisme.

[31] La directive intitulée "Vie privée et communications électroniques" s'applique aux courriers électroniques et aux messages textes. Le nouveau règlement en la matière s'applique de manière plus large; il porte sur les données créées ou traitées par les nouvelles formes de communication électronique, notamment les communications de machine à machine, la téléphonie Internet et les services d'accès à l'Internet.

des factures de services publics ou des frais de scolarité peut indiquer un flux de trésorerie régulier et une approche positive quant au paiement des dettes.

L'accès à ces données s'avère un bon moyen d'initier aux services financiers numériques des personnes qui en étaient exclues jusqu'à présent – notamment en raison du manque d'informations à leur sujet. Dans de nombreux cas, les opérateurs de réseaux mobiles se sont associés à des banques pour faciliter les prêts mobiles. Cette démarche s'appuie sur les scores de crédit définis à partir des données relatives à leurs clients. L'opérateur ne partage pas toujours avec les banques les données brutes relatives à l'argent mobile ni les métadonnées relatives aux appels, mais il y appliquera souvent des algorithmes pour produire un score de crédit.

Par exemple[32], un opérateur de réseau mobile utilise 48 paramètres sur une période de six mois, ainsi que les informations recueillies dans le cadre du processus d'enregistrement des personnes (connaissance des clients) pour produire un tableau de bord et définir plusieurs catégories de clients présentant des caractéristiques similaires. Au-dessus des scores de crédit se trouvent des "règles commerciales" fixées par la banque qui déterminent les limites réelles qui peuvent être proposées aux clients. Il peut notamment s'agir d'un seuil à atteindre ou à maintenir pour pouvoir prétendre à un prêt, de plafonds s'appliquant à différentes catégories de clients ou de limites individuelles (par exemple, des limites de crédit déterminées par une formule appliquée à la moyenne mensuelle des rentrées d'argent mobile d'un client), mais aussi d'obstacles à l'entrée (par exemple, clients figurant sur une liste noire en raison d'un défaut de paiement passé ou d'un fichier de crédit négatif dans une agence d'évaluation du crédit).

Les entreprises qui proposent des services de paiement à valeur ajoutée sont également de plus en plus à même d'utiliser les données relatives aux paiements pour prendre des décisions en matière de crédit. Par exemple, la société Kopo Kopo facilite l'accès des commerçants au système de paiement M-Pesa de Safaricom, au Kenya, grâce à des interfaces de programmation d'applications (API) qui permettent aux commerçants de recevoir des paiements et qui assurent la gestion du processus de réception ainsi que la comptabilité des reçus de paiement. L'entreprise dispose ainsi d'un aperçu unique des flux de trésorerie de ses clients commerçants, et elle est idéalement placée pour évaluer leur solvabilité et développer ainsi une activité de prêt.

Données sur les activités en ligne

Au-delà de l'utilisation par un client des services d'un opérateur de réseau mobile, de grandes quantités de données provenant d'activités de navigation en ligne et d'applications de téléphonie mobile sont recueillies et partagées, souvent sans faire l'objet de politiques d'acceptation expresse standard. Par exemple, une récente étude de l'Université d'Oxford portant sur environ un million d'applications Android a révélé que près de 90% d'entre elles communiquaient des informations à Google[33]. Les informations relatives à l'utilisation d'Internet peuvent être recueillies en même temps que les données de localisation, les informations de contact et les messages textes (voir figure 2).

---

[32] Cet exemple est tiré d'un accord de coopération conclu entre un grand opérateur de réseau mobile et une banque de renom. Ledit accord est disponible auprès de l'auteur. Les noms demeurent confidentiels.

[33] Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T. et Shadbolt, N., *Third Party Tracking in the Mobile Ecosystem*, arXiv:1804.03603v3 [cs.CY], 18 octobre 2018, disponible à l'adresse suivante: https://arxiv.org/pdf/1804.03603.pdf; Ram, A., Wisniewska, A., Kao, J. S., Rininsland, Æ. et Nevitt, C., "How smartphone apps track users and share data", *Financial Times*, 23 octobre 2018, disponible à l'adresse suivante: https://ig.ft.com/mobile-app-data-trackers/.

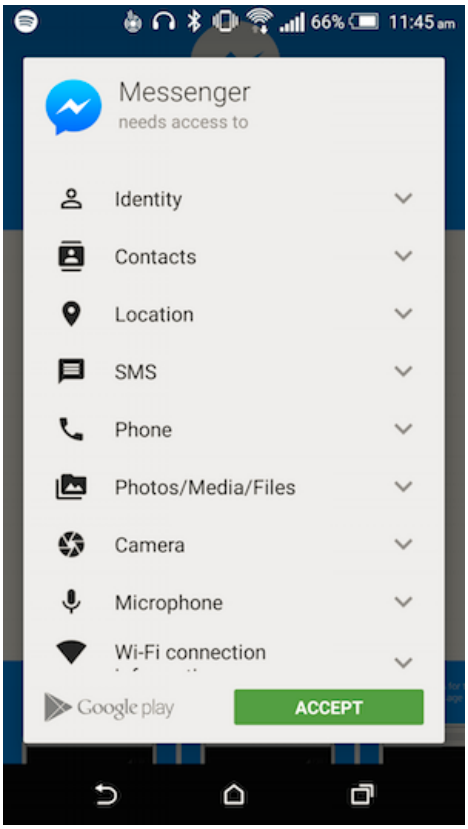

Figure 2 – Paramètres d'autorisation des applications pour smartphones

Le marché des données rend possible un suivi en ligne ainsi qu'un suivi multiappareils, qui permet de relier les données d'utilisation d'un smartphone à celles d'un ordinateur et d'une tablette, par exemple. Avec le développement de l'Internet des objets, les données provenant des appareils qu'une personne utilise au travail, chez elle ou sur son corps seront de plus en plus interconnectées. En raison de ce large éventail de sources de données interconnectées, il est possible de suivre les déplacements d'un utilisateur grâce à des applications de cartographie, à l'historique de son navigateur et de ses recherches, à ce qu'il "aime" et qui il "aime" sur les réseaux sociaux, aux vidéos ou à la musique qu'il a visionnées ou écoutée en streaming, à l'historique de ses achats, au contenu de ses articles de blog et de ses commentaires publiés en ligne, et bien plus encore. Des entreprises telles que Branch, Tala et Jumo ont développé d'importantes activités de crédit numérique en Afrique en s'appuyant sur ces données alternatives.

Types de données générales

Il existe de nombreuses autres sources de données relatives aux personnes qui peuvent être combinées pour les besoins d'opérations ayant recours aux mégadonnées. Ces données peuvent être recueillies auprès des commerces de détail où une personne effectue des achats ou des sociétés de cartes de crédit utilisées pour les transactions, ou bien à partir de dispositifs de détection Bluetooth installés dans les magasins (collecte de données passive), de photos de personnes prises par caméra, de plaques d'immatriculation de voitures enregistrées par caméra, d'informations sur les médicaments recueillies lors d'achats de produits pharmaceutiques, ainsi que d'enregistrements effectués par des jouets dotés d'un microphone ou d'une caméra, pour ne citer que quelques sources. Un conseiller auprès d'investisseurs

actifs sur le marché des mégadonnées recense les sources suivantes de données alternatives disponibles sur le marché actuel[34]:

- données provenant d'agrégateurs financiers;
- données de cartes de crédit;
- données géospatiales et de localisation;
- ensembles de données recueillies sur Internet;
- données sur l'utilisation des applications;
- données des douanes américaines relatives aux expéditions;
- données sur les dépenses publicitaires;
- données mises à disposition sur des API;
- données de localisation/circulation piétonnière provenant de capteurs et de routeurs;
- données provenant des médias sociaux;
- données interentreprises acquises auprès des parties prenantes de la chaîne d'approvisionnement;
- données sur l'agriculture (par exemple, les aliments provenant de la production de maïs);
- données provenant des points de vente;
- données sur les prescriptions pharmaceutiques.

La connectivité croissante des appareils offre aux fournisseurs de services financiers la possibilité d'utiliser certaines données. Par exemple, les voitures d'aujourd'hui disposent d'une puissance de calcul considérable, utilisent un codage avancé et traitent d'énormes volumes de données[35]. Avant d'octroyer un prêt, les prêteurs exigent de plus en plus souvent des emprunteurs, en particulier des emprunteurs à haut risque, qu'ils consentent à l'installation de dispositifs de suivi dans leur voiture (dispositifs d'interruption de démarrage, par exemple). Les dispositifs d'interruption de démarrage ont l'avantage de contribuer à l'application des droits de reprise de possession, car ils permettent au prêteur de désactiver un véhicule si l'emprunteur ne rembourse pas son prêt. Dans le même temps, ces dispositifs de suivi recueillent des données sur les activités courantes et les lieux visités au quotidien, permettant ainsi de déterminer l'adresse du domicile et du lieu de travail, si la personne conduit toujours pour se rendre à son lieu de travail habituel (ce qui permet donc également de déduire sa situation professionnelle), où la personne aime faire des achats ou se divertir, et tout écart par rapport aux habitudes, ce qui peut indiquer des changements de préférences. Les dispositifs de suivi peuvent également fournir des données sur le comportement des personnes au volant, qui indiquent leur niveau de compétence, voire même parfois un état émotionnel ou psychologique particulier (par exemple, des accélérations répétées et inhabituellement rapides, ou un freinage inhabituellement brusque).

Aujourd'hui, déduire des informations sur les individus constitue un marché important. La manière dont elles sont générées et utilisées est abordée dans la section suivante. De manière générale, la relation entre l'intelligence artificielle et les mégadonnées est "bidirectionnelle". Les mégadonnées s'appuient sur

[34] ZwillGen, "Alternative Data: Best Practices", présenté au Forum sur la vie privée et la sécurité, à Washington D. C., 2018.
[35] En 2014, McKinsey estimait que la voiture de l'époque avait la puissance de calcul de 20 ordinateurs de bureau, nécessitait environ 100 millions de lignes de code de programmation et traitait jusqu'à 25 gigaoctets de données par heure. McKinsey, "What's Driving the Connected Car". Septembre 2014. Disponible à l'adresse suivante: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/whats-driving-the-connected-car.

l'intelligence artificielle et l'apprentissage automatique pour extraire de la valeur de grands ensembles de données, et l'apprentissage automatique repose sur un volume énorme de données[36].

4.3 Qu'entend-on par profilage et décisions automatisées?

Les mégadonnées, l'apprentissage automatique et l'intelligence artificielle offrent des débouchés commerciaux rentables et des avantages sociaux grâce au ***profilage*** et aux ***décisions automatisées***.

Le ***profilage*** désigne le traitement automatisé de données à caractère personnel visant à évaluer, analyser ou prédire divers aspects probables des intérêts d'une personne, de ses préférences personnelles, de son comportement, de sa productivité au travail, de sa situation économique, de sa santé, de sa fiabilité, de sa localisation ou de ses déplacements[37]. L'analyse des données permet d'identifier les liens qui existent entre les individus et de définir des profils de groupe[38].

Ces déductions et prédictions peuvent être utilisées dans le cadre de publicités ciblées ou pour prendre des ***décisions automatisées*** (ou étayer des décisions humaines). Les décisions automatisées sont des décisions prises par des systèmes de traitement informatique sans aucune implication humaine (au-delà du codage) et généralement fondées sur des déductions faites à partir du profilage, à l'aide de modèles d'apprentissage automatique appliqués aux mégadonnées. Les déductions et les prédictions améliorent la capacité des entreprises à différencier les consommateurs, en leur proposant des produits et des services adaptés à leurs préférences ou à leurs besoins, et à des prix qu'ils sont prêts à payer. Elles étayent par exemple les décisions d'accorder un crédit à une personne ou de lui proposer un emploi.

De nombreuses applications des mégadonnées et de l'apprentissage automatique sont introduites dans les services financiers, notamment:

- l'évaluation des risques, que ce soit pour des prêts ou des assurances (comme expliqué précédemment), par des sociétés telles que Compare.com[39];
- les "robots conseillers" pour la gestion de portefeuilles d'investissement, tels que Betterment[40] et Wealthfront[41], qui s'appuient sur des algorithmes pour calibrer un portefeuille financier en fonction des objectifs d'investissement et de la tolérance au risque de l'usager;
- la négociation à haute fréquence par des fonds spéculatifs et d'autres institutions financières, telles que Walnut Algorithms[42] et Renaissance Technologies[43], qui utilisent l'apprentissage automatique pour prendre des décisions en matière de négociation en temps réel[44];

---

[36] 38e Conférence Internationale des commissaires à la protection des données et de la vie privée, *Artificial Intelligence, Robotics, Privacy and Data Protection*, 2016. Disponble à l'adresse suivante: https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf.

[37] Le paragraphe 4 de l'article 4 du RGPD définit le "profilage" 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".

[38]

[39] https://www.compare.com/.

[40] https://www.betterment.com/.

[41] https://www.wealthfront.com/.

[42] https://walnut.ai/en/.

[43] https://www.rentec.com/Home.action?index=true.

[44] https://www.quora.com/Why-are-machine-learning-neural-networks-and-other-AI-approaches-for-instance-not-more-widely-used-in-stock-market-predictions.

- la gestion d'actifs, l'évaluation du risque de liquidité et de change, ainsi que les tests de résistance;
- la détection de fraudes par des sociétés comme APEX Analytics[45] et Kount[46], grâce à la détection et au signalement d'activités inédites ou d'anomalies de comportement, afin de bloquer les transactions et de permettre aux équipes de sécurité d'enquêter; et
- une multitude de services tels que la sécurité et l'identification numérique, l'analyse de l'actualité, les ventes et les recommandations aux consommateurs, et le service client[47].

Dans certains cas, ces nouvelles utilisations sont encouragées par une législation autorisant expressément le recours à l'intelligence artificielle. Au Mexique, par exemple, les réformes de 2018 s'appliquant aux fintech ont modifié la loi relative au marché de valeurs mobilières afin d'adopter des règles spéciales régissant les services de conseil et de gestion des investissements automatisés (également connus sous le nom de robots conseillers)[48].

4.4 En quoi consiste la protection des usagers?

Il s'agit d'une démarche visant à protéger les êtres humains là où ils sont vulnérables. Il peut s'agir de la protection des enfants, des personnes âgées et d'autres individus qui ne peuvent se protéger eux-mêmes pour des raisons physiques ou psychologiques. Il est toutefois généralement reconnu que tous les usagers sont vulnérables à certains égards. Nous ne pouvons pas toujours tout savoir. Notre capacité à évaluer les risques et les avantages est limitée – nous sommes soumis à une "rationalité limitée"[49].

La protection des usagers implique l'intervention de l'État, qui met en place des lois et des procédures régissant ce qui serait autrement une relation privée entre l'usager et le fournisseur. Cette nécessité découle des asymétries perçues entre les fournisseurs et les usagers. Il peut s'agir d'asymétries en matière d'information, lorsque les fournisseurs disposent davantage de données et de connaissances, ainsi que d'une meilleure compréhension que les usagers. Les différences d'échelle économique peuvent également entraîner d'importantes asymétries dans le pouvoir de négociation. En outre, les coûts de transaction auxquels les consommateurs seraient confrontés s'ils devaient négocier des assurances pour chaque

[45] https://www.apexanalytix.com/.
[46] https://www.kount.com/.
[47] Faggella, D., "Machine Learning in Finance – Present and Future Applications". 18 septembre 2018. Disponible à l'adresse suivante: https://www.techemergence.com/machine-learning-in-finance/.
[48] Article 227 bis 1 de la loi mexicaine relative au marché de valeurs mobilières, chapitre sur les conseillers en investissement.
[49] Voir, par exemple: Sunstein, C., Jolls, C. et Thaler, R., "A Behavioural Approach to Law and Economics". *Stanford Law Review*, vol. 50, 1998; et Sunstein, C. et Thaler, R., *Nudge*, Yale University Press, 2008.

produit ou service acheté sont trop élevés pour être envisageables. Par conséquent, une relation exclusivement privée et négociée entre le consommateur et le fournisseur constituerait un marché unilatéral.

Si elle est formulée de diverses manières, la protection des usagers cherche généralement à promouvoir les valeurs d'équité, de responsabilité et de transparence[50]. Le débat politique autour de la protection des usagers en ce qui concerne l'intelligence artificielle et l'apprentissage automatique se concentre sur la capacité des algorithmes et des systèmes d'apprentissage automatique à refléter de telles valeurs[51]. Les usagers peuvent être vulnérables lorsqu'ils ont affaire à des services reposant sur le traitement informatique, et ce, pour de nombreuses raisons. Leur fonctionnement dépasse l'entendement de la majorité de la population. Leurs savants processus numériques et leurs résultats possèdent une "séduisante précision de production"[52]. Les ordinateurs et les résultats qu'ils produisent peuvent donc être perçus comme des éléments objectifs, voire équitables. Aujourd'hui, les usagers risquent toutefois de considérer certains aspects des services numériques comme injustes, non responsables et manquant de transparence (à l'opposé des principes d'équité, de responsabilité et de transparence), une situation qui sape la confiance envers les fournisseurs de services, entravant ainsi les perspectives de croissance des services numériques.

Les lois relatives à la protection des consommateurs reposent généralement sur l'application de règles, de principes et de procédures visant à donner aux consommateurs certains droits relatifs aux produits et services qu'ils achètent. Il s'agit notamment:

- de droits antérieurs à l'achat (***pré-engagement***), tels que les informations sur le produit ou le service proposé;
- de la fourniture, de la qualité et du fonctionnement du produit ou du service lui-même (***engagement***); et
- des moyens de demander des comptes aux fournisseurs après l'achat (***post-engagement***).

Les principes d'équité, de responsabilité et de transparence peuvent s'appliquer à l'étape de pré-engagement, lorsqu'on exige que les consommateurs soient informés du produit ou du service qu'ils achètent et que l'on cherche éventuellement à obtenir leur consentement explicite, afin qu'ils puissent assumer la responsabilité de leurs décisions.

Toutefois, une grande partie de la législation sur la protection des consommateurs part du principe que, même si le consommateur est informé de l'existence d'un produit ou d'un service et qu'il y consent conformément aux conditions générales connexes, ce consentement à lui seul peut ne pas suffire à garantir l'application des principes d'équité, de responsabilité et de transparence. Ainsi, ces principes peuvent également s'appliquer à l'étape d'engagement, c'est-à-dire au produit ou au service lui-même, à savoir sa sûreté, sa qualité ou d'autres caractéristiques et conditions de fourniture. Par conséquent, les

---

[50] Par exemple, la section 5 de la loi sur la Commission fédérale du commerce des États-Unis interdit les actes ou pratiques déloyaux ou trompeurs dans le cadre du commerce, ou qui touchent au commerce, ce qui fut l'un des fondements de l'application de la législation sur la protection de la vie privée numérique aux États-Unis.

[51] Crawford, K., *et al*., *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term.* 7 juillet 2016, p. 6-8. Disponible à l'adresse suivante: https://artificialintelligencenow.com/media/documents/AINowSummaryReport_3_RpmwKHu.pdf.

[52] Schwartz, P., "Data Processing and Government Administration: The Failure of the American Legal Response to the Computer". *Hastings Law Journal*, vol. 43, 1992, p. 1321 et 1342.

lois relatives à la protection des consommateurs vont au-delà des obligations d'information et de consentement préalables à l'engagement lorsque ces démarches ne protègent pas suffisamment le consommateur, et ne doivent pas alléger la responsabilité du fournisseur.

Les principes d'équité, de responsabilité et de transparence s'appliquent également à l'étape de post-engagement, afin de garantir la mise en place de mécanismes de redevabilité permettant d'obtenir des explications sur les raisons pour lesquelles un produit ou un service donné a été fourni de telle ou telle manière. Ils donnent aux consommateurs la possibilité de contester ces décisions et constituent une voie de recours en cas de préjudice. Ces protections peuvent être appliquées indépendamment du fait que le consommateur ait donné son consentement ou non. Par exemple, les lois de nombreux pays ne permettent pas aux consommateurs de se soumettre à certains types de procédures d'arbitrage en cas de réclamation et de renoncer à leur droit d'être entendus par un tribunal. Au contraire, elles insistent sur les procédures qui garantissent aux consommateurs un processus équitable et transparent pour demander des comptes aux fournisseurs.

Ces lois protègent ainsi les consommateurs contre les descriptions de produits trompeuses, les clauses contractuelles abusives (par exemple, l'exclusion de toute responsabilité), les produits défectueux et l'absence de dispositifs de recours. Ces lois interdisent aux fabricants et aux détaillants de négocier de telles conditions avec les consommateurs, de sorte qu'ils ne peuvent pas prétendre que ces derniers y ont consenti lorsqu'ils ont acheté tel ou tel produit ou service. La stratégie de protection des consommateurs introduit des normes et des procédures communes minimales en vue d'assurer un niveau de protection de base, plutôt que d'engager l'autonomie et la responsabilité des consommateurs.

Les lois relatives à la protection des consommateurs ont un lien étroit, voire symbiotique, avec le droit et les politiques en matière de concurrence. L'asymétrie du pouvoir de négociation qui justifie la protection des consommateurs peut être exacerbée lorsqu'un marché est concentré et que les consommateurs manquent d'alternatives pour un service donné. S'agissant des politiques en matière de concurrence, on cherche de plus en plus à s'attaquer aux niveaux élevés de concentration sur les marchés de données. La Commission européenne et plusieurs États membres ont élaboré différentes théories du préjudice autour des grandes entreprises technologiques qui recueillent des données sur les consommateurs par l'intermédiaire de modèles commerciaux, qui utilisent ces données pour générer des revenus publicitaires. Certaines autorités, telles que le Bundeskartellamt, l'autorité allemande de la concurrence, ont évoqué la possibilité que le non-respect du droit à la vie privée des consommateurs puisse, dans certaines circonstances, constituer un abus de position dominante sur le marché dans le cadre du droit de la concurrence. Le présent document ne porte toutefois pas sur les aspects des mégadonnées et de l'apprentissage automatique liés au droit de la concurrence, mais sur les questions relatives à la protection des consommateurs et de la vie privée.

Un certain nombre de mesures de protection des consommateurs examinées dans le présent document sont tout aussi pertinentes pour les entreprises individuelles que pour les micro, petites et moyennes entreprises. Lorsque la législation nationale ne les considère pas comme des sujets de données ou des consommateurs, ces entreprises peuvent ne pas bénéficier des protections offertes par les lois sur la protection des données et de la vie privée. Il existe de solides arguments en faveur de l'extension de ces protections à ces entreprises.

4.5 Qu'entend-on par confidentialité des données?

Risques pour la vie privée

Toutes les techniques liées aux mégadonnées et à l'apprentissage automatique ne reposent pas sur des données personnelles et n'entraînent pas de problèmes en matière de protection des usagers. Il existe de nombreuses données qui ne concernent aucune personne identifiable et qui peuvent être utilisées à des fins commerciales et sociales. Toutefois, lorsque des données personnelles sont utilisées, elles peuvent susciter des inquiétudes quant à la vie privée des personnes concernées.

La vie privée englobe un large éventail de notions. Qu'elle soit considérée comme une valeur ou en termes de droits ou de protections, elle a été réduite par certains chercheurs à des préoccupations concernant l'individualité, l'autonomie, l'intégrité et la dignité[53], qui font partie d'un éventail plus large d'idées centrées sur la liberté dans la vie personnelle et familiale.

Si la vie privée peut désigner l'absence d'interférence avec les choix personnels des individus, notamment pour ce qui est de leur corps, une grande partie de cette notion concerne ce que certaines personnes connaissent à propos de quelqu'un, et donc le traitement des données personnelles. Confidentialité des données et sécurité des données ne sont pas la même chose. La gestion sécurisée des données est nécessaire pour protéger la vie privée, mais cette dernière concerne des valeurs spécifiques relatives à chaque personne qui doivent être prises en compte pour garantir la sécurité des données.

Ainsi, dans un contexte numérique, la protection de la vie privée implique des mesures de contrôle pour la collecte, l'utilisation et le partage des données personnelles. L'expression "données à caractère personnel" a une signification potentiellement vaste, qui s'étend à toute information relative à un individu identifiable[54]. La plupart des dispositifs de protection des données reconnaissent que certaines données personnelles sont plus sensibles ou plus facilement sujettes à des abus que d'autres, et appliquent des contrôles renforcés en conséquence.

Les données concernant une personne peuvent avoir été:

- communiquées par la personne (par exemple, un nom d'utilisateur ou un code postal);
- observées (par exemple, les données de localisation); ou
- obtenues à partir d'informations communiquées ou observées (par exemple, le pays de résidence déduit en fonction du code postal); ou
- déterminées à partir de ce qui précède (par exemple, un score de crédit) par déduction ou raisonnement logique[55].

Les usagers sont exposés à des risques d'atteinte à la vie privée lorsque leurs données personnelles peuvent i) être consultées par des personnes qui n'en ont pas le droit, ii) faire l'objet d'abus ou iii) être

---

[53] Bygrave, L. A., "Data Protection Law: Approaching Its Rationale, Logic and Limits". *Kluwer Law International*, 2002, p. 128-129.
[54] Le RGPD définit les "données à caractère personnel" au paragraphe 1 de l'article 4 comme "toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée 'personne concernée'); est réputée être une 'personne physique identifiable' une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale".
[55] Le groupe de travail "Article 29" de l'Union européenne sur la protection des données a établi une distinction entre ces trois catégories dans les "Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679" (n 19) 8. Disponible à l'adresse suivante: http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

utilisées à des fins de profilage, ce qui conduit à des déductions subjectives à propos des usagers, lesquelles peuvent être difficiles à vérifier, et aboutir à des décisions automatisées qui affectent la vie des personnes concernées.

L'un des principaux risques pour la vie privée est lié à l'agrégation des données personnelles. Dans le cas des mégadonnées, ce risque est exacerbé lorsque les données personnelles ne sont pas anonymisées, ou que des tentatives de pseudonymisation ou d'anonymisation ont été effectuées, mais que la réidentification des personnes concernées reste possible (voir section 6.3). De plus en plus, les pays légifèrent pour protéger les données personnelles et la vie privée des personnes concernées, une question importante étant la minimisation de la collecte, de l'utilisation et du partage des données.

Avec les mégadonnées et l'apprentissage automatique, les données à caractère personnel susceptibles d'être générées et partagées peuvent inclure des déductions faites au sujet des personnes concernées et des prédictions sur leur comportement. Les déductions qui ont été faites à propos d'une personne à partir de ses données personnelles ne sont généralement pas considérées comme des données personnelles à protéger[56]. La législation réduit souvent la protection de la vie privée à la rectification, au blocage ou à l'effacement des données personnelles introduites dans les algorithmes, sans tenir compte de l'évaluation de ces données ou des décisions qui s'appuient dessus. Comme cela a récemment été suggéré à propos du RGPD, ironiquement, de tous les types de données abordés dans la législation sur la protection des données, ce sont les déductions faites à partir de données personnelles qui sont le moins protégées, alors qu'elles posent peut-être aujourd'hui les plus grands risques pour la vie privée et la lutte contre la discrimination[57].

Protection de la vie privée

Les recours possibles en matière de protection des données comprennent le droit des usagers de savoir quelles données personnelles sont recueillies[58], le droit de rectifier des données personnelles inexactes et de compléter des données personnelles incomplètes[59], le droit de faire supprimer des données personnelles[60], le droit de transférer des données à une partie tierce[61], et le droit de s'opposer au traitement de données personnelles (y compris à des fins de profilage)[62]. Si l'Union européenne a adopté l'ensemble de ces recours dans le RGPD, de nombreux pays se concentrent davantage sur les droits d'accès et de rectification, ainsi que sur les obligations de signalement des violations.

La protection des données et le respect de la vie privée ne sont pas l'apanage des pays à revenu élevé de l'hémisphère Nord. Aujourd'hui, 107 pays, dont 66 pays en développement ou en transition, ont adopté des lois sur la protection des données et de la vie privée, et d'autres sont en passe de le faire[63]. De

---

[56] Wachter, S. et Mittelstadt, B., "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI". *Columbia Business Law Review*, 2019. Disponible à l'adresse suivante: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248829.
[57] *Ibid.*
[58] RGPD, articles 13 à 15.
[59] RGPD, article 16.
[60] RGPD, article 17.
[61] RGPD, article 20.
[62] RGPD, article 21.
[63] Conférence des Nations Unies sur le commerce et le développement, Global cyberlaw tracker, en date du 27 septembre 2018. À la suite d'une nouvelle mesure, ce nombre est passé à 120 en 2017. Voir: Greenleaf, G., "Global Data Privacy Laws 2017: 120 National Data Privacy Laws, Including Indonesia and Turkey", *Privacy Laws & Business International Report*, vol. 145, document de recherche nº 17-45 de UNSW Law, 30 janvier 2017, p. 10-13. Rien qu'en Afrique, 22 pays disposent déjà de lois sur la protection de la vie privée et des données: l'Afrique du Sud (2013), l'Angola (2016), le Bénin (2009), le Botswana (2018), le Burkina Faso (2004), Cabo Verde (2001), la

nombreux pays non européens se sont engagés à respecter des niveaux stricts de protection des données en signant la Convention 108 (par exemple, le Mexique l'a signée en 2018).

Le RGPD de l'Union européenne ne prévoit pas seulement des droits et des obligations renforcés, il a également une grande incidence extraterritoriale. Ce document exige que les données personnelles soient protégées lorsqu'elles sont exportées et traitées dans des pays non européens. Il s'applique au traitement des données concernant toute personne qui se trouve "dans l'Union", même si celui-ci a lieu en dehors de l'Union européenne. Ainsi, les pays traitant avec l'Europe dans le cadre de services numériques et les entreprises non européennes susceptibles d'utiliser les données des Européens doivent adopter des mesures de protection comparables à celles du RGPD. Par exemple, le Japon a clôturé les discussions visant à établir des systèmes de protection des données et de la vie privée suffisamment proches de ceux de l'Union européenne pour mériter l'étiquette "en adéquation" en 2018, et des échanges sont en cours avec la Corée du Sud. Le système de l'Uruguay a été jugé adéquat en 2012 en vertu de la directive européenne antérieure sur la protection des données.

Certains pays considèrent la protection des données et de la vie privée comme une question de droit constitutionnel. La Constitution mexicaine, par exemple, interdit toute intrusion dans la personne, la famille, le domicile, les documents ou les biens d'un individu (y compris toute mise sur écoute de dispositifs de communication), sauf sur ordre d'une autorité compétente soutenue par le droit applicable[64]. Le droit à la protection des données est prévu par la loi; il fixe une norme concernant la collecte, l'utilisation, la conservation, la divulgation ou le transfert (collectivement dénommés le traitement) de données personnelles, afin de garantir le droit à la vie privée et à l'autodétermination[65].

En 2017, la Cour suprême de l'Inde a déclaré que la vie privée était un "droit fondamental", protégé par la Constitution[66], rejoignant ainsi les États-Unis[67], l'Union européenne[68] et de nombreuses autres juridictions. Dans certains cas, ces questions sont explicitement inscrites dans la Constitution elle-même. La Constitution brésilienne, par exemple, prévoit un droit d'"*habeas data*" qui autorise les personnes à accéder aux données personnelles les concernant qui sont détenues par des organismes publics, et à les corriger[69]. Certains pays, comme le Kenya, prévoient un droit constitutionnel à la vie privée, mais n'ont pas (encore) introduit de législation autonome.

La prolifération des données et la capacité des technologies ayant recours aux mégadonnées à porter atteinte à la vie privée ont récemment conduit la Cour suprême indienne à limiter l'utilisation d'Aadhaar,

---

Côte d'Ivoire (2013), le Gabon (2011), le Ghana (2012), la Guinée équatoriale (2016), le Lesotho (2012), Madagascar (2014), le Mali (2013), Maurice (2017), la Mauritanie (2017), le Maroc (2009), le Sénégal (2008), les Seychelles (2002), le Tchad (2015), la Tunisie (2004), la Zambie et le Zimbabwe (2003). L'Algérie, l'Éthiopie, le Kenya, le Malawi, la Mauritanie, le Niger, le Nigéria, l'Ouganda, la République démocratique du Congo, le Rwanda, la Sierra Leone, le Swaziland et la Tanzanie ont quant à eux préparé des projets de loi. Voir: Collaboration sur la politique internationale des TIC pour l'Afrique orientale et australe (CIPESA), *State of Internet Freedom in Africa 2018 – Privacy and Data Protection in the Digital Era: Challenges and Trends in Africa*, septembre 2018, p. 7.

[64] Paragraphes 1 et 12 de l'article 16 de la Constitution mexicaine.

[65] Paragraphe 2 de l'article 16 de la Constitution mexicaine.

[66] Juge K. S. Puttaswamy et Anr. c. Union of India et Ors, 10 SCC 1, 2017.

[67] En 1974, le Congrès américain a stipulé dans le Privacy Act, une loi fédérale sur la protection de la vie privée, que le droit à la vie privée était un droit personnel et fondamental protégé par la Constitution des États-Unis.

[68] En décembre 2009, lors de l'entrée en vigueur du Traité de Lisbonne, la Charte des droits fondamentaux de l'Union européenne garantissait la protection de la vie privée et des données parmi 50 autres droits fondamentaux.

[69] Article 5 (LXXII), Constitution de la République fédérative du Brésil, 3e édition, 2010. Disponible à l'adresse suivante: http://english.tse.jus.br/arquivos/federal-constitution.

le système national d'identification numérique du pays[70]. La Cour a déclaré que le fait d'exiger l'utilisation d'Aadhaar pour des services autres que les services publics (par exemple, les allocations sociales), y compris à des fins de connaissance des clients au sein du secteur bancaire ou des télécommunications, serait désormais illégal[71]. Elle estime que les conditions légales visant spécifiquement à relier le système Aadhaar à l'ensemble des comptes bancaires et numéros de téléphone mobile, qu'ils soient nouveaux ou non, violaient le droit fondamental à la vie privée, jugeant que cela permettrait l'exploitation commerciale des informations biométriques et démographiques d'un individu par des entités privées.

Faire de la vie privée un droit fondamental n'est qu'une stratégie parmi d'autres pour assurer la protection des usagers. Certains pays considèrent la vie privée moins comme un droit fondamental que comme une question liée à la protection des usagers. Bien que ce point de vue puisse se traduire par un moindre engagement en faveur de la protection générale de la vie privée, cela peut permettre de se concentrer davantage sur les compromis et les questions de rentabilité liées à la réglementation de la protection de la vie privée. Les organismes de protection des usagers devront plus souvent chercher à trouver un juste équilibre lorsqu'il s'agira de déterminer si un comportement donné est injuste pour les usagers et doit être considéré comme illégal[72].

Cette approche n'empêche pas de centrer la législation et la réglementation sur la protection de la vie privée là où elle est la plus importante, ce qui, dans la plupart des pays, inclut les secteurs de la santé, des finances et des communications, ainsi que la protection de l'enfance. Certains pays ne disposent pas de loi d'application générale sur la protection de la vie privée, mais ont élaboré une législation et une réglementation substantielles en la matière dans les différents secteurs, à des moments différents et sans harmonisation poussée avec les dispositions juridiques sectorielles. Si cette démarche peut permettre d'adapter les préoccupations en matière de protection de la vie privée aux spécificités d'un secteur donné, cela risque également de créer des difficultés, des incohérences entre les secteurs et des problèmes d'harmonisation entre les pays.

Certains pays ont préféré établir des normes non contraignantes pour la protection de la vie privée, comme les normes nationales chinoises régissant les technologies de sécurité des données personnelles ("Information Security Technology – Personal Information Security Specification GB/T 35273-2017") entrées en vigueur en 2018. Ledit document définit de nombreuses normes pour la protection des données personnelles, vaguement inspirées du RGPD européen. Il définit en outre les pratiques que les organismes de régulation s'attendront à voir introduites lorsqu'ils auditeront les entreprises et appliqueront les lois chinoises en vigueur sur la protection des données, en particulier la loi de 2016 relative à la cybersécurité. D'autres normes nationales, notamment sur les mégadonnées et l'anonymisation des données, devraient être introduites.

Même les juridictions qui stipulent que la vie privée est un droit fondamental reconnaissent la nécessité de trouver un équilibre entre l'intérêt de l'individu et celui des organisations publiques et privées, ainsi

---

[70] Pour de plus amples informations sur Aadhaar, voir https://uidai.gov.in/.
[71] Juge K. S. Puttaswamy et Anr. c. Union of India et Ors, paragraphes 159-160, 2018. Disponible à l'adresse suivante: https://www.sci.gov.in/supremecourt/2012/35071/35071_2012_Judgement_26-Sep-2018.pdf.
[72] La Commission fédérale du commerce des États-Unis, l'organisme général de réglementation de la protection de la vie privée, est soumise à un équilibre statutaire: la Commission n'est pas habilitée à déclarer illégal un acte ou une pratique au motif que cet acte ou cette pratique sont déloyaux, à moins qu'ils ne causent ou ne soient susceptibles de causer aux usagers un préjudice substantiel qui ne peut être raisonnablement évité par les usagers eux-mêmes et qui n'est pas contrebalancé par des avantages compensatoires ni par la concurrence (titre 15, paragraphe 45 n du Code des États-Unis).

que les intérêts sociaux plus larges, tels que la recherche scientifique, l'innovation, la sécurité nationale et la lutte contre la criminalité. Le droit fondamental de mener une activité commerciale existe dans de nombreuses juridictions[73], tout comme les droits de propriété intellectuelle et les secrets commerciaux.

La protection de la vie privée, comme toute réglementation, implique des coûts, tels que les coûts financiers de la mise en conformité et les coûts d'opportunité des nouveaux services reposant sur l'accès aux données personnelles. Certains affirment que ces coûts constituent un investissement économique justifiable, car le renforcement de la confiance augmentera la demande de services. D'autres considèrent ces investissements comme un choix du type de société dans laquelle nous voulons vivre, pour reprendre les propos de Tim Cook, directeur général d'Apple[74].

Dans tous les cas, il est raisonnable et approprié pour les organismes de législation et de régulation de prendre en compte non seulement l'idéal de la vie privée, mais aussi les obstacles à l'innovation et aux objectifs productifs, ainsi que le détournement de ressources, que les mesures de protection axées sur la conformité peuvent entraîner. Il est prudent d'identifier et de quantifier au mieux les avantages et les coûts, et d'accorder la priorité aux risques les plus dangereux. Comme le soulignent la Banque mondiale et le Groupe consultatif pour l'aide aux plus pauvres (CGAP)[75], les décideurs politiques doivent relever le défi de trouver un juste équilibre entre la promotion des avantages de l'utilisation élargie des données alternatives et la garantie d'une protection adéquate des données et d'une prise en compte de la vie privée des usagers dans l'ensemble de l'écosystème[76].

5 La phase de pré-engagement: difficultés liées à la protection des usagers et de la vie privée en matière d'information et de consentement

Cette section s'intéresse à l'obligation, prévue par de nombreuses lois sur la protection des usagers et de la vie privée, d'informer les usagers du fait que leurs données personnelles seront recueillies, utilisées et partagées avec des parties tierces, ainsi que de la finalité de ces processus, et d'obtenir leur consentement – et ce, avant qu'ils ne s'engagent à communiquer des données et à solliciter un service.

5.1 Informer les usagers et obtenir leur consentement avant d'utiliser leurs données personnelles

Les lois et normes relatives à la protection des données d'un nombre croissant de pays prévoient une stricte réglementation de la collecte, de l'utilisation et du partage des données. Celles-ci exigent des entreprises qu'elles avertissent les usagers lorsqu'elles recueillent des données à caractère personnel les concernant, qu'elles les informent de la finalité du traitement de ces données et qu'elles leur demandent leur consentement avant de communiquer ces données à de tierces parties[77]. Ces dernières peuvent

[73] Voir par exemple l'article 16 de la Charte des droits fondamentaux de l'Union européenne.
[74] Discours prononcé à l'occasion de la Conférence internationale des commissaires à la protection des données et de la vie privée, à Bruxelles, le 24 octobre 2018. Transcription complète disponible à l'adresse suivante: https://www.computerworld.com/article/3315623/security/complete-transcript-video-of-apple-ceo-tim-cooks-eu-privacy-speech.html.
[75] Le CGAP est une branche de la Banque mondiale qui se concentre sur la réduction de la pauvreté par l'inclusion financière. Voir https://www.cgap.org/about/governance.
[76] Banque mondiale et CGAP, "Data Protection and Privacy for Alternative Data", document de travail provisoire du sous-groupe du GPFI sur la loi relative à la protection financière des consommateurs (*Financial Consumer Protection Law*). 4 mai 2018, p. 5.
[77] RGPD, article 13. La norme chinoise sur la sécurité des informations personnelles de 2018 exige que les personnes concernées soient informées de la portée, de la finalité et des principes du traitement de leurs informations personnelles de manière explicite, compréhensible et raisonnable.

également être tenues d'informer les usagers lorsqu'elles obtiennent des informations personnelles les concernant[78]. Cependant, cette démarche est rarement obligatoire, et même lorsque c'est le cas, elle peut se limiter à certaines catégories d'informations et non aux déductions concernant les personnes.

Deux questions de longue date intéressant la législation sur la protection des données et de la vie privée sont la "spécification des finalités" et, partant, la "minimisation des données", à savoir, respectivement, i) l'obligation de préciser la finalité pour laquelle les données sont recueillies, utilisées et partagées, et ii) l'obligation de limiter la collecte, l'utilisation et le partage aux données qui sont pertinentes, adéquates et nécessaires à cette finalité (ou proportionnelles à celle-ci)[79]. Comme toute collecte et utilisation de données peut augmenter les risques liés à la sécurité et à la confidentialité, l'objectif est de limiter ou d'éviter tout risque superflu par rapport à ce qui est strictement nécessaire aux fins définies. Il s'agit d'éviter tout "détournement d'usage", c'est-à-dire l'utilisation de données à d'autres fins que celles définies à l'origine[80]. Le "Principe de la limitation de l'utilisation" de l'OCDE, par exemple, fait référence à la nécessité d'obtenir le consentement des personnes si les données les concernant doivent être utilisées à d'autres fins que celles pour lesquelles elles ont été initialement recueillies[81].

Il existe parfois des exceptions aux règles en matière d'information et de consentement qui permettent d'utiliser les données au-delà de leur objectif initial de collecte, par exemple à des fins statistiques ou de recherche scientifique[82], qui nécessitent souvent de grands ensembles de données, pour la même raison que l'apprentissage automatique en général. Il existe parfois des zones grises entre ce qui constitue des objectifs statistiques ou de recherche scientifique et ce qui relève de la mise au point de produits dans le cadre de la prestation de services financiers. Toutefois, le champ d'application de ces exceptions aux règles de spécification des finalités et de minimisation des données n'est généralement pas très important.

[78] RGPD, article 14.
[79] Par exemple, la norme chinoise sur la sécurité des informations personnelles de 2018 prévoit que, sauf accord contraire des personnes concernées, les responsables du contrôle des données personnelles doivent limiter ce processus aux démarches nécessaires pour atteindre l'objectif défini et supprimer ces informations une fois ledit objectif mené à bien.
[80] Voir plus généralement: Kindt, E. J., *Privacy and Data Protection: Issues of Biometric Application, A Comparative Analysis.* Springer, Heidelberg, Dordrecht, New York, London, 2013.
[81] OCDE, "Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel", telles que modifiées en 2013, principe 10. Disponible à l'adresse suivante: https://www.oecd.org/fr/sti/ieconomie/lignesdirectricesregissantlaprotectiondelaviepriveeetlesfluxtransfrontieresdedonneesdecaractereperso nnel.htm.
[82] RGPD, paragraphe 50 du préambule et alinéa 1b de l'article 5.

Principes FEAT de l'Autorité monétaire de Singapour

2. L'utilisation de caractéristiques personnelles comme facteurs d'entrée pour prendre des décisions grâce aux technologies d'intelligence artificielle et d'analyse des données (de l'anglais *Artificial Intelligence and Data Analytics*, ou AIDA) est justifiée.

12. Pour accroître la confiance du public, les personnes concernées sont informées de manière proactive de l'utilisation des technologies d'AIDA dans le cadre d'une communication générale.

13. Les personnes concernées qui en font la demande reçoivent des explications claires sur les données utilisées pour prendre des décisions les concernant grâce aux technologies d'AIDA et sur la manière dont ces données influent sur lesdites décisions.

14. Les personnes concernées qui en font la demande reçoivent des explications claires sur les conséquences que les décisions prises grâce aux technologies d'AIDA peuvent avoir pour elles.

Version préliminaire des normes de la Smart Campaign relatives au crédit numérique

Indicateur 6.1.1.1

Le fournisseur a évalué et documenté les informations personnelles dont il a besoin de la part de ses clients afin d'assurer le service proposé (par exemple, identité, transactions, etc.). La collecte, le partage et la durée de conservation des données personnelles sont réduits au strict nécessaire et directement justifiés par la loi. L'évaluation a permis d'identifier les risques que la collecte, le traitement, la conservation et la communication des données personnelles représentent pour la vie privée des consommateurs.

Indicateur 6.1.1.6

Les données à caractère personnel doivent être i) pertinentes au regard des finalités pour lesquelles elles seront utilisées et, dans la mesure nécessaire à ces fins, ii) exactes, complètes et à jour.

Indicateur 6.2.1.0

Les personnes concernées sont invitées à consentir aux utilisations qui seront faites de leurs données personnelles. Les demandes de consentement expliquent clairement, dans un langage simple et dans la langue locale, la manière dont les données seront utilisées. Un consentement distinct est requis pour: a) le partage des données avec certaines parties tierces (à identifier clairement) dans le cadre de la prestation de services; b) la communication des données aux agences d'évaluation du crédit; c) l'utilisation des données à des fins de marketing; d) la vente de données à des parties tierces; et e) l'utilisation des données de géolocalisation. S'agissant des services reposant sur des données de service complémentaire non structurées (USSD) ou sur des messages textes, fournir les hyperliens vers les accords de divulgation ne suffit pas.

Indicateur 6.2.2.0

Le droit de renoncer à un service et de retirer l'autorisation accordée à une organisation d'utiliser des données (de quelque type que ce soit) doit être clairement indiqué et accessible aux consommateurs, de même que les conséquences d'un tel retrait.

Indicateur 6.2.3.0

Les consommateurs ont le droit d'obtenir du fournisseur la confirmation que celui-ci possède ou non des données les concernant, et si cette demande est rejetée, ils ont le droit de savoir pourquoi.

Indicateur 6.2.3.1

Les consommateurs ont le droit de se voir communiquer les données les concernant dans un délai raisonnable, sans frais excessifs et en des termes qu'ils peuvent comprendre.

Les lois de nombreux pays, ainsi que les normes internationales et régionales, exigent également que les personnes donnent leur accord pour la collecte, l'utilisation et le partage de leurs données personnelles[83]. Lorsqu'un tel accord n'est ni exigé ni obtenu, certaines juridictions permettent aux usagers de "se retirer" en indiquant qu'ils ne souhaitent pas que leurs données personnelles soient recueillies, utilisées ou partagées avec de tierces parties[84]. Lorsque les consommateurs n'ont pas le choix, les lois sur la protection des données peuvent imposer des obligations de transparence, exigeant des responsables du contrôle des données qu'ils fournissent des explications claires et accessibles dans les politiques de confidentialité quant à la manière dont leurs données seront utilisées et partagées et à quelles fins[85].

En ce qui concerne les décisions prises dans le cadre des mégadonnées et de l'apprentissage automatique, une approche consiste simplement à les proscrire lorsqu'elles présentent un risque inacceptable. Cette démarche a notamment été recommandée pour l'utilisation d'armes létales. À quelques exceptions près, les voitures automatisées ne sont pas encore autorisées sur les routes, bien que des lois soient en cours d'élaboration pour faire évoluer les choses.

Cependant, compte tenu des avantages de nombreux processus automatisés pour les usagers, ceux-ci sont souvent autorisés, à condition d'informer les usagers de toute prise de décisions automatisée et de leur donner la possibilité de s'y opposer. Par exemple, le RGPD exige notamment d'informer les personnes concernées de "l'existence d'une prise de décision automatisée, y compris un profilage [...] et, au moins en pareil cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée"[86]. Il prévoit également, au paragraphe 1 de l'article 22, que "la personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire"[87].

S'il peut s'avérer utile, ce droit de retrait est limité. Les décisions automatisées sont autorisées par le RGPD lorsqu'elles sont nécessaires à la conclusion d'un contrat avec la personne concernée, ou avec son consentement[88]. Lorsque de nouveaux services, qui sont censés être fournis rapidement, souvent à distance et par voie électronique, s'appuient sur le profilage pour définir l'éligibilité d'une personne, des décisions automatisées peuvent être nécessaires pour conclure le contrat. Et lorsque le besoin ou le

[83] Par exemple, le RGPD de 2016 indique dans son préambule, au paragraphe 40: "Pour être licite, le traitement de données à caractère personnel devrait être fondé sur le consentement de la personne concernée ou reposer sur tout autre fondement légitime prévu par la loi, soit dans le présent règlement soit dans une autre disposition du droit national ou du droit de l'Union [...]." On entend par consentement "toute manifestation de volonté, libre, spécifique, éclairée et univoque par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement" (alinéa 11 de l'article 4 du RGPD).

[84] Par exemple, *California Consumer Privacy Act* (loi de la Californie sur la protection de la vie privée des consommateurs), 2018.

[85] Par exemple, dans son *Cadre de protection de la vie privée* de 2004, la Coopération économique Asie-Pacifique (APEC) exige que les responsables du contrôle des données fournissent des explications claires et facilement accessibles sur leurs pratiques et leurs politiques en matière de protection des informations personnelles.

[86] RGPD, articles 13, 14 et 15.

[87] De même, le projet de loi sur la protection des données au Kenya, qui devrait bientôt entrer en vigueur, prévoit à l'article 31 que toute personne a le droit de ne pas faire l'objet d'une décision fondée uniquement sur un traitement automatisé de ses données personnelles, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative. Il prévoit également des exceptions, notamment si le traitement automatisé s'avère nécessaire à l'exécution d'un contrat, qu'il est autorisé par la loi et s'accompagne de garanties, et qu'il repose sur un consentement explicite.

[88] RGPD, article 22.

souhait d'obtenir un produit ou un service dépasse l'aversion à faire l'objet d'un traitement automatisé, un choix binaire s'offre à la personne concernée, qui peut n'avoir d'autre option que de consentir.

5.2 Tenir compte des difficultés posées par les mégadonnées

Les mégadonnées et l'apprentissage automatique entravent les démarches d'information et de consentement prévues par la législation et la réglementation sur la protection des données et de la vie privée.

Spécification des finalités dans le cadre de l'apprentissage automatique

Se conformer aux exigences d'information implique d'indiquer en détail aux personnes concernées la finalité de la collecte de leurs données personnelles, et de surveiller étroitement les opérations pour éviter de dépasser cet objectif. L'apprentissage automatique décèle des formes, puis creuse dans des couches plus profondes, identifiant d'autres formes, qui peuvent révéler des cas d'utilisation sans rapport direct avec l'objectif initial de l'exploration des données. De ce fait, il est possible que la finalité de l'utilisation des données ne soit pas clairement définie et que seuls de vagues objectifs aient été identifiés au moment de la collecte des données ou de l'obtention du consentement, ce qui explique d'ailleurs la nature généralement incertaine des politiques de confidentialité et des notifications de collecte de données.

La minimisation des données dans le contexte des mégadonnées

En outre, les techniques d'apprentissage automatique étant plus efficaces pour détecter des formes à partir d'ensembles de données élargis au fil du temps, la nature même des mégadonnées est de recueillir le plus grand nombre possible de données – ainsi que de les conserver le plus longtemps possible.

Ainsi, la notion même de minimisation des données (recueillir aussi peu de données que possible et les conserver le moins longtemps possible, en fonction de l'objectif défini) va à l'encontre du *modus operandi* de l'industrie. Elle compromet la possibilité d'informer avec précision les personnes concernées de l'objet de la collecte de leurs données personnelles. Les processus de divulgation, de contrôle et de mise en conformité peuvent également s'avérer coûteux et difficiles à mener à bien. Définir un objectif très large afin d'éviter de telles limites pourrait bien ne pas être légalement acceptable. Comme l'indique un rapport remis au Président des États-Unis en 2014, l'information et le consentement sont compromis par les avantages mêmes offerts par les mégadonnées: des méthodes d'utilisation nouvelles, imprévues et étonnamment puissantes des données[89].

Les limites de la responsabilité des usagers

En outre, malgré les efforts déployés pour rendre les informations simples et compréhensibles, il est rare que les usagers lisent et comprennent ces documents[90]. Cette situation compromet les démarches d'information et de consentement, les rendant non seulement inefficaces, mais aussi trompeuses; elle fait

---

[89] Conseil présidentiel des conseillers en science et technologie, *Big Data and Privacy: A Technological Perspective.* Maison-Blanche, Washington, D. C., 1er mai 2014.

[90] Voir, par exemple: Whitley, E. A. et Pujadas, R., *Report on a study of how consumers currently consent to share their financial data with a third party*. Financial Services Consumer Panel, p. ii, 2018. Les résultats des études empiriques indiquent que le consentement n'est souvent ni librement donné, ni sans ambiguïté, ni tout à fait éclairé. Plus de la moitié des personnes interrogées ont déclaré ne pas lire les conditions générales des produits et services auxquels elles souscrivent, y compris les services qui ont accès à leurs données financières. De même, seule une petite proportion de participants a répondu correctement à une question sur un point de la politique, même après avoir eu l'occasion de relire le document dans le cadre de la recherche.

souvent peser sur l'usager un fardeau qu'il n'est pas en mesure d'assumer et crée une impression de légitimité injustifiée. Les politiques de confidentialité et la procédure de consentement peuvent répondre au strict minimum dans le cadre d'une approche axée sur la conformité, mais elles ne permettent guère aux usagers de comprendre comment leurs données peuvent être utilisées et partagées avec des parties tierces, sans parler des implications qui en découlent[91].

Certains ont suggéré de simplifier les informations, parce que les spécifications de conception de l'intelligence artificielle et de l'apprentissage automatique ne peuvent garantir une responsabilité et une vérifiabilité satisfaisantes, et de les rendre plus percutantes – à l'image de la tête de mort trouvée sur les produits de nettoyage ménagers qui contiennent des composés toxiques[92].

Outre le fait qu'il est difficile d'attendre des usagers qu'ils assument la responsabilité de questions qui dépassent souvent leur entendement, la manière dont le consentement est sollicité, à savoir sur une base binaire "à prendre ou à laisser", ne fait qu'exacerber le problème.

La confidentialité en contexte

Certains suggèrent que l'une des approches consiste à reconnaître que la confidentialité est généralement très étroitement liée au contexte, ainsi qu'aux attentes qu'une personne pourrait raisonnablement avoir à la lumière de la nature de la situation ou de la transaction. Une personne peut s'attendre à un niveau élevé de confidentialité (traitement confidentiel) sur des questions personnelles (médicales, financières, etc.), mais pas lorsqu'on surprend l'une de ses conversations dans un endroit public ou qu'on lui propose de l'aider à trouver certains produits dans un magasin. Dans le cadre d'une étude, les attentes en matière de confidentialité peuvent varier en fonction du contexte, notamment de l'objet ou de l'objectif de l'étude en question. De même, la question de savoir si l'on peut s'attendre à pouvoir profiter d'un divertissement en privé peut dépendre de la nature du contenu.

Il a donc été suggéré que le contexte, et non l'économie politique, devait déterminer les contraintes relatives aux flux d'informations, de sorte que les mesures de protection de la vie privée en ligne correspondent à ces attentes[93]. Cela pourrait impliquer des restrictions plus sévères en matière de collecte, d'utilisation et de partage des données personnelles dans certaines situations, même lorsque des procédures d'information et de consentement sont en place. La Charte des droits des consommateurs en matière de protection de la vie privée (*Consumer Privacy Bill of Rights*) proposée par la Maison-Blanche du président Obama en 2012 [94] cherchait à s'aligner sur cette stratégie, adoptant comme troisième principe le "respect du contexte", soit le fait de s'attendre à ce que les entreprises recueillent, utilisent et divulguent les données personnelles en fonction du contexte dans lequel ces informations leur ont été communiquées[95].

Dans la mesure où le consentement de l'utilisateur continue d'être considéré comme une base légitime pour la collecte et l'utilisation des données personnelles, les procédures de consentement pourraient être

---

[91] Solove, D. J., "Privacy Self-Management and the Consent Dilemma". *Harvard Law Review*, vol 126, nº 1880, 2013, p. 1889-1893.
[92] Initiative mondiale de l'IEEE (voir la note de bas de page 223), p. 159.
[93] Nissenbaum, H., *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Stanford University Press, 2010; et Nissenbaum, H., *A Contextual Approach to Privacy Online*, Daedalus, 2011, disponible à l'adresse suivante: https://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf.
[94] http://btlj.org/2012/03/president-obamas-privacy-bill-of-rights-encouraging-a-collaborative-process-for-digital-privacy-reform/.
[95] *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy*, 2012. Disponible à l'adresse suivante: https://obamawhitehouse.archives.gov/sites/default/files/privacy-final.pdf.

améliorées. Outre la simplification du langage dans lequel sont communiquées les informations, ces améliorations peuvent inclure le recours à un consentement à plusieurs niveaux qui différencie plusieurs types de données en fonction des différents types d'objectifs pour lesquels elles peuvent être utilisées ou des différents types d'organisations qui peuvent les utiliser. Des dispositions de temporisation prévoyant l'expiration du consentement peuvent également s'avérer appropriées[96].

Technologies de gestion du consentement

Des efforts sont également déployés pour mettre au point des technologies et des services permettant de mieux gérer le consentement. Cette démarche repose sur plusieurs aspects: employer différentes formes de gestion des droits numériques, joindre des autorisations aux données personnelles et permettre des négociations automatisées, entre les personnes qui communiquent leurs données personnelles et celles qui les reçoivent, concernant les processus de collecte, d'utilisation et de partage. Ces approches visent à améliorer la transparence et le contrôle des données par les usagers, et donc à faciliter leur accessibilité grâce à une confiance accrue[97]. Au lieu de décisions de consentement binaires, dans le cadre desquelles les usagers autorisent l'accès à toutes leurs données, sans quoi ils ne peuvent pas profiter du service recherché, il peut y avoir des moyens de permettre un consentement gradué en fonction des préférences des usagers en matière de partage et de conservation de leurs données personnelles.

Pour que cela soit possible à grande échelle, il faudra peut-être faire appel à des outils algorithmiques agissant en tant qu'agent[98], tuteur ou fiduciaire – des "anges gardiens algorithmiques"[99] – au nom de l'usager. Certains suggèrent que les fournisseurs de services de gestion des données personnelles de ce type pourraient informer et éduquer chaque usager et procéder à des "négociations" en son nom, en lui montrant comment les données demandées pourraient être combinées avec d'autres données fournies précédemment, en l'avertissant de toute utilisation non autorisée de ses données personnelles, ou en lui prodiguant des conseils en fonction de son profil[100]. Ce processus pourrait même impliquer la définition de conditions relatives au partage des données: rémunération des usagers, droit de revenir sur un consentement antérieur lorsque les conditions n'ont pas été respectées, etc.

Il semble exister une véritable opportunité commerciale d'investissement et d'innovation dans l'amélioration de la gestion du consentement des usagers. Des entreprises comme Sudo[101] permettent aux usagers d'utiliser facilement un pseudonyme pour diverses interactions numériques, des appels téléphoniques au commerce électronique ou aux rencontres en ligne. Apple prévoit d'introduire une fonction de connexion anonyme dans les applications mobiles, qui utilise des adresses électroniques

---

[96] Voir, par exemple: Custers, B., *Click Here to Consent Forever: Expiry Dates for Informed Consent.* Big Data & Society, janvier-juin 2016. Disponible à l'adresse suivante: http://journals.sagepub.com/doi/10.1177/2053951715624935.
[97] Pentland, A., "Big Data's Biggest Obstacles". *Harvard Business Review*, 2012. 2012. Disponible à l'adresse suivante: https://hbr.org/2012/10/big-datas-biggest-obstacles.
[98] Une norme s'appliquant aux "acteurs de l'intelligence artificielle" est en cours d'élaboration dans le cadre du projet de l'IEEE *"P7006 – Standard for Personal Data Artificial Intelligence (AI) Agent*". Disponible à l'adresse suivante: https://standards.ieee.org/project/7006.html.
[99] Koponen, J. M., "We need algorithmic angels". TechCrunch, 2014. Disponible à l'adresse suivante: https://techcrunch.com/2015/04/18/we-need-algorithmic-angels/.
[100] Voir *Ethically Aligned Design*, à la note de bas de page 223, p. 103. Voir également: Orcutt, M., "Personal AI Privacy Watchdog Could Help You Regain Control of Your Data", *MIT Technology Review*, 11 mai 2017, disponible à l'adresse suivante: https://www.technologyreview.com/s/607830/personal-ai-privacy-watchdog-could-help-you-regain-control-of-your-data/; et l'application mobile "Privacy Assistant" correspondante, disponible à l'adresse suivante: https://play.google.com/store/apps/details?id=edu.cmu.mcom.ppa&hl=en.
[101] https://mysudo.com/.

générées de manière aléatoire. Il s'agirait d'une alternative aux applications qui proposent des inscriptions par le biais de comptes de médias sociaux tiers comme Facebook, afin de réduire la dépendance à l'égard des fournisseurs qui suivent les utilisateurs et vendent des publicités en fonction de leurs habitudes[102].

Des idées similaires impliquent généralement un plus grand contrôle des usagers sur leurs données personnelles. Par exemple, le "casier numérique" de l'Inde, qui fait partie de l'India Stack, permet aux individus d'avoir un plus grand contrôle sur les personnes qui peuvent accéder à leurs données, notamment en créant un registre vérifiable des accès à leurs dossiers. La définition d'un droit de propriété sur les données personnelles est une autre idée, bien que cette tendance n'ait pas encore pris de l'ampleur.

Toutes ces suggestions visent à renforcer le contrôle des usagers sur leurs données personnelles, en réduisant les asymétries qui prévalent actuellement. La qualité des données recueillies peut même s'en trouver améliorée. Certains suggèrent que le fait de permettre aux personnes de définir un certain degré d'anonymat lorsqu'elles répondent à des demandes de collecte de données (par exemple, après un achat ou dans le cadre d'une enquête de santé) peut améliorer la fiabilité des données communiquées[103].

6 La phase d'engagement: protection des usagers et de la vie privée dans le cadre des services ayant recours à l'intelligence artificielle

Cette section s'intéresse à l'engagement des usagers: leur expérience avec les mégadonnées et l'apprentissage automatique, et inversement la collecte, l'utilisation, la conservation et le transfert de leurs données personnelles par les entreprises utilisant les mégadonnées et l'apprentissage automatique. Les sections 6.1 et 6.2 étudient les préoccupations des usagers et les problèmes juridiques qui découlent des résultats substantiels obtenus à partir du traitement des données, en particulier la responsabilité en matière d'exactitude et la prise de décisions biaisées. La section 6.3 porte sur les mesures de protection des usagers contre le risque de divulgation de leurs données personnelles en cas de réidentification et de non-respect de la confidentialité des données, en se concentrant sur les techniques de désidentification, de pseudonymisation et d'anonymisation. La section 1.1 aborde les risques pour les usagers qui découlent des transferts de données sur le marché dynamique des courtiers en données, et de la réglementation accrue de ce secteur de marché.

6.1 Précision – protéger les usagers contre les données erronées et obsolètes

Exactitude des données d'entrée

Le bon fonctionnement des modèles d'apprentissage automatique et la précision de leurs résultats dépendent de l'exactitude des données d'entrée. Certains des vastes volumes de données utilisés pour

---

[102] Herrera, S. et Haggin, P., "New Apple Sign-In Option Could Keep More Personal Data Away From Facebook, Google". *The Wall Street Journal*, 6 juin 2019. Disponible à l'adresse suivante: https://www.wsj.com/articles/new-apple-sign-in-option-could-keep-more-personal-data-away-from-facebook-google-11559839438.

[103] Wong, K.-S. et Kim, M. H., "Towards a respondent-preferred k_i-anonymity model". *Frontiers of Information Technology & Electronic Engineering*, vol. 16, 2015, p. 720. Disponible à l'adresse suivante: https://doi.org/10.1631/FITEE.1400395. Le niveau d'anonymat (c'est-à-dire le *k*-anonymat) garanti par un organisme ne peut pas être vérifié par les personnes concernées puisqu'elles n'ont généralement pas accès à l'ensemble des données qui sont publiées. Par conséquent, nous introduisons la notion de k_i-anonymat, où *k* est le niveau d'anonymat préféré par chaque personne interrogée *i*. Au lieu de faire entièrement confiance à un organisme, notre solution renforce la confiance des individus en permettant à chacun de décider du niveau de protection qu'il préfère. En tant que tel, notre protocole garantit le degré k_i-anonymat choisi par les personnes interrogées pendant la collecte des données ainsi que l'authenticité et l'utilité des informations recueillies aux fins de l'analyse des données.

entraîner le système peuvent être "structurés" (organisés et facilement consultables) et d'autres, "non structurés"[104]. Les données peuvent avoir été obtenues de différentes manières au fil du temps, à partir de plusieurs sources, et plus ou moins directement. Plus l'ensemble de données recueillies est grand, plus il y a de chances que les données soient obsolètes et que les processus de mise à jour systématique ne soient pas appliqués. Les données antérieures peuvent même avoir été incorrectes depuis le départ.

Ces facteurs peuvent compromettre l'exactitude des données alimentant les algorithmes. Cela vaut à la fois pour les données à caractère personnel d'un individu qui fait l'objet d'une décision automatisée (auxquelles on applique le modèle d'apprentissage automatique), ainsi que pour l'ensemble de données élargi utilisé pour entraîner l'ordinateur. Si les données d'apprentissage sont inexactes, le modèle ne permettra pas de produire les résultats escomptés lorsqu'il sera appliqué aux données personnelles. Tous ces problèmes peuvent donner lieu à des déductions erronées sur les usagers.

Les lois sur la protection des données et de la vie privée imposent donc de plus en plus aux entreprises une certaine forme d'obligation légale visant à garantir l'exactitude des données qu'elles détiennent et manipulent. La législation mexicaine en matière de protection des données applique un principe de qualité exigeant des responsables du contrôle des données qu'ils vérifient que les données à caractère personnel figurant dans leurs bases de données sont correctes, à jour et correspondent aux fins pour lesquelles elles ont été recueillies[105].

Cela soulève la question de l'exactitude des données dans l'écosystème des données au sens large, et de la mesure dans laquelle les entreprises devraient être tenues responsables de leur inexactitude ou contribuer à améliorer la qualité des informations de manière générale.

Responsabilité en matière d'exactitude des données dans les services financiers

Les lois sectorielles régissant les services financiers soulignent souvent l'importance de garantir l'exactitude des données utilisées dans le cadre de ces services. Les données utilisées aux fins de l'évaluation du crédit en sont un exemple[106]. Les agences d'évaluation du crédit sont généralement soumises à une réglementation et à des contrôles internes stricts afin de garantir l'exactitude des données qu'elles détiennent sur les particuliers. Ces systèmes d'évaluation du crédit réduisent le coût des

---

[104] Les données structurées présentent un degré élevé d'organisation, de sorte que leur inclusion dans une base de données relationnelle est transparente et facilement consultable par de simples algorithmes de moteur de recherche ou d'autres opérations de recherche (par exemple, comptes rendus de paiements et de transactions). Les données non structurées ne suivent pas de modèle prédéfini et ne sont pas organisées de manière prédéfinie (par exemple, les entrées de médias sociaux, les courriels et les images).

[105] De la même manière, l'article 5 (paragraphe 1d) du RGPD stipule que les données à caractère personnel doivent être "exactes et, si nécessaire, tenues à jour; [et que] toutes les mesures raisonnables doivent être prises pour que les données à caractère personnel qui sont inexactes, eu égard aux finalités pour lesquelles elles sont traitées, soient effacées ou rectifiées sans tarder".

[106] Les principes relatifs à la qualité des données de l'OCDE (principe 8), le Cadre de protection de la vie privée de l'APEC (principe 21), le principe relatif à la qualité des données de la Résolution de Madrid et la Convention 108 (article 5) comportent tous des dispositions exigeant que les informations soient exactes et à jour. Les Principes de haut niveau du G20 sur l'inclusion financière numérique appellent à l'élaboration de directives visant à garantir l'exactitude et la sécurité de toutes les données relatives aux comptes et aux transactions; au marketing des services financiers numériques; et à l'élaboration de scores de crédit pour les consommateurs financièrement exclus et mal desservis. Ces orientations devraient couvrir les formes traditionnelles et nouvelles de données, telles que les données sur le paiement des factures de services publics, sur l'achat de temps de communication mobile, sur l'utilisation de portefeuilles numériques ou de comptes d'argent électronique, sur les médias sociaux et sur les transactions électroniques (https://www.gpfi.org/sites/gpfi/files/documents/G20%20High%20Level%20Principles%20for%20Digital%20Financial%20Inclusion%20-%20Full%20version-.pdf).

opérations de prêt en diminuant le risque inhérent aux asymétries d'information entre prêteurs et emprunteurs (et donc les pertes sur prêts, les provisions pour créances douteuses et le besoin de garanties). Ils fournissent aux prêteurs des informations qui leur permettent d'évaluer les emprunteurs, favorisant ainsi un meilleur accès aux services financiers[107]. En raison de l'importance de leurs données dans la prise de décisions en matière de crédit et autres, les agences d'évaluation du crédit donnent aux particuliers les moyens de corriger les informations inexactes.

Toutefois, ce système d'information formel n'est plus qu'une partie d'un écosystème élargi riche en données, dont la plupart ne sont pas réglementées. L'avènement des mégadonnées et de l'apprentissage automatique présente le risque que la législation et les orientations politiques existantes ne suivent pas le rythme auquel évolue cet écosystème. Par exemple, le premier des Principes généraux sur l'évaluation du crédit de la Banque mondiale (de l'anglais *General Principles on Credit Reporting*) publiés en 2011[108] stipule que les systèmes d'évaluation du crédit doivent i) disposer de données pertinentes, exactes, à jour et suffisantes – ainsi que positives – recueillies de manière systématique auprès de toutes les sources fiables, adéquates et disponibles, et ii) conserver ces informations suffisamment longtemps.

On pourrait se poser des questions sur la manière dont ce type d'orientations politiques devraient s'appliquer aujourd'hui – à peine huit ans plus tard – aux données personnelles communiquées et recueillies à des fins qui n'étaient peut-être pas initialement liées à la prise de décisions en matière de crédit. Les mégadonnées et l'apprentissage automatique peuvent conduire à collecter et à utiliser des données dont la pertinence, l'exactitude et le degré d'actualité varient considérablement.

Ces difficultés concernent également les lois qui ont été rédigées avant l'avènement des mégadonnées et de l'apprentissage automatique, voire d'Internet lui-même. Les entreprises qui ne se considèrent pas comme des agences d'évaluation du crédit peuvent toutefois être soumises aux obligations légales qui s'appliquent aux agences d'évaluation du crédit traditionnelles. Dans certains cas, ces entreprises pourraient faire l'objet de réclamations pour avoir communiqué des informations inexactes ayant influencé la solvabilité perçue d'une personne.

De nombreux pays reconnaissent qu'il est d'intérêt public de garantir "des évaluations du crédit justes et précises", comme cela a été formulé aux États-Unis, par exemple[109]. Cela permet à la fois de favoriser le fonctionnement des marchés de services financiers et de protéger les consommateurs. Pour cette raison, les agences de renseignement sur les consommateurs dont les données sont utilisées dans plusieurs domaines (transactions de crédit, assurances, octroi de licences, transactions commerciales initiées par les consommateurs, emploi, etc.) sont souvent réglementées[110].

Cependant, les lois de nombreux pays relatives aux renseignements sur les consommateurs ont été promulguées avant l'avènement d'Internet, sans parler des mégadonnées et de l'apprentissage automatique. Certains pays se font une idée plus générale des agences de renseignement sur les consommateurs. Aux États-Unis, par exemple, la loi sur les rapports de crédit équitables (en anglais *Fair Credit Reporting Act*, ou FCRA) s'applique aux entreprises qui diffusent régulièrement des informations

---

[107] Société financière internationale, *Credit reporting knowledge guide*, Washington, D. C., 2012. Disponible à l'adresse suivante: https://www.ifc.org/wps/wcm/connect/2bc067fb-80e8-429c-b2e1-2e6e059fc153/Credit+Reporting+lowres+NEW.pdf?MOD=AJPERES&CVID=jMq.GVR.

[108] http://www.worldbank.org/en/topic/financialsector/publication/general-principles-for-credit-reporting.

[109] Paragraphe 1681(a)(1).

[110] Paragraphes 1681a(d)(1)(A)-(C) et 1681b.

sur la solvabilité, la capacité de crédit, la réputation générale, les caractéristiques personnelles ou le mode de vie d'une personne[111]. Ladite loi exige des agences de renseignement sur les consommateurs qu'elles suivent des procédures raisonnables pour garantir la plus grande exactitude possible des rapports sur les consommateurs; qu'elles informent les fournisseurs et les utilisateurs d'informations sur les consommateurs des responsabilités qui leur incombent en vertu de la loi; qu'elles limitent les circonstances dans lesquelles ces acteurs fournissent des rapports sur les consommateurs "à des fins d'emploi"; et qu'elles publient des numéros gratuits permettant aux consommateurs d'obtenir lesdits rapports. Elle engage également la responsabilité des parties prenantes en cas de non-respect de ces exigences[112].

Dans un rapport de 2016, l'autorité américaine de protection des consommateurs, la Commission fédérale du commerce, s'est penchée sur la manière dont les mégadonnées étaient utilisées dans les décisions découlant des évaluations du crédit[113]. Celle-ci a précisé que les courtiers en données qui compilaient des informations non traditionnelles, y compris des informations provenant des médias sociaux pouvaient être considérés comme des agences d'évaluation du crédit, et partant, être soumis aux mêmes obligations.

Il ne s'agit pas d'une simple possibilité théorique. Par exemple, dans une récente affaire de la Cour suprême des États-Unis opposant Spokeo et Robins[114], Spokeo exploitait un site Internet visant à rechercher et à recueillir des données à partir d'un large éventail de bases de données. Spokeo renseignait l'adresse, le numéro de téléphone, l'état civil, l'âge approximatif, la profession, les loisirs, la situation financière, les habitudes d'achat et les préférences musicales des individus, et permettait aux utilisateurs de rechercher des informations sur d'autres personnes. Le plaignant, Robins, a avancé que Spokeo l'avait décrit de manière incorrecte comme un professionnel riche et marié, ce qui le rendait surqualifié pour certains emplois convoités. Robins a affirmé que Spokeo était une agence de renseignement sur les consommateurs au sens du FCRA[115], et que la société était donc responsable de la communication d'informations incorrectes. L'affaire a été résolue sur d'autres motifs, mais l'envergure potentielle de cette législation héritée pose des problèmes aux entreprises menant des activités dans le domaine des données. Cette situation peut engager la responsabilité des parties prenantes envers les consommateurs quant à l'exactitude des données utilisées pour prendre des décisions en matière de crédit et autres qui n'ont pas

---

[111] Paragraphe 1681a(d)(1).
[112] Paragraphes 1681e(b); 1681e(d); 1681b(b)(1); 1681j(a); et 1681n(a).
[113] Commission fédérale du commerce des États-Unis, *Big data: A tool for inclusion or exclusion? Understanding the issues*. Washington, D. C., 2016. Disponible à l'adresse suivante: https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.
[114] 136 S. Ct. 1540, 2016.
[115] 15e Congrès des États-Unis, "Fair Credit Reporting Act". *Statutes at Large*, vol. 84, tel que modifié, 1970, p. 1127, paragraphe 1681 et paragraphes suivants. Le FCRA vise à garantir l'établissement de rapports de crédit équitables et exacts (paragraphe 1681(a)(1)). Il réglemente l'établissement et l'utilisation de rapports sur les consommateurs par les agences de renseignement sur les consommateurs dans des domaines tels que les transactions de crédit, les assurances, l'octroi de licences, les transactions commerciales initiées par les consommateurs et l'emploi (paragraphes 1681a(d)(1)(A)-(C) et 1681b). Le FCRA a été promulgué bien avant l'avènement d'Internet et s'applique aux entreprises qui diffusent régulièrement des informations portant sur la solvabilité, la capacité de crédit, la réputation générale, les caractéristiques personnelles ou le mode de vie d'un individu (paragraphe 1681a(d)(1)). Il exige des agences de renseignement sur les consommateurs qu'elles suivent des procédures raisonnables pour garantir l'exactitude maximale des rapports sur les consommateurs; qu'elles informent les fournisseurs et les utilisateurs d'informations sur les consommateurs des responsabilités qui leur incombent en vertu de la loi; qu'elles limitent les circonstances dans lesquelles ces acteurs fournissent des rapports sur les consommateurs à des fins d'emploi; et qu'elles publient des numéros gratuits permettant aux consommateurs d'obtenir lesdits rapports. Il engage également la responsabilité des parties prenantes en cas de non-respect de ces exigences (paragraphes 1681e(b); 1681e(d); 1681b(b)(1); 1681j(a); et 1681n(a)).

été anticipées, affaiblir la sécurité juridique et compromettre l'innovation et les investissements en faveur des entreprises.

Les exigences en matière d'évaluation du crédit et l'écosystème d'information au sens large

La section ci-dessus portait sur les responsabilités envers les consommateurs que les entreprises peuvent avoir lorsqu'elles traitent des données de manière non traditionnelle, en particulier en ce qui concerne l'exactitude des données qu'elles utilisent pour prendre des décisions dans le domaine des services financiers. Une question connexe se pose concernant la responsabilité des entreprises de contribuer à l'écosystème d'information élargi qui est traditionnellement réglementé par des obligations de divulgation et d'information.

Les obligations de divulgation apparaissent dans de nombreux contextes, qu'il s'agisse des exigences légales relatives aux valeurs mobilières qui s'appliquent aux sociétés publiques, des informations sur la santé et la sécurité des médicaments, ou de produits de consommation qui présentent des risques particuliers. Dans le contexte des services financiers, par exemple, les antécédents de crédit d'une personne constituent des données utiles pour un fournisseur de services financiers et réduisent l'asymétrie d'information entre le prêteur et l'emprunteur. Afin d'améliorer la concurrence entre les fournisseurs de services qui détiennent ces données et le fonctionnement des marchés financiers, certains fournisseurs de services financiers sont souvent tenus de communiquer les données relatives au crédit des consommateurs à des agences de renseignement sur les consommateurs qui les organisent et les mettent à la disposition du marché dans son ensemble.

Dans de nombreux pays, seules les banques (c'est-à-dire les entités qui sont réglementées, généralement par des licences bancaires, pour la réception de dépôts, l'octroi de prêts et d'autres activités connexes) sont tenues de faire rapport aux agences d'évaluation du crédit afin qu'elles intègrent les informations dont elles disposent à leurs dossiers et leurs analyses. Aujourd'hui, la question est de savoir si les fournisseurs de services financiers non bancaires qui s'appuient sur des décisions automatisées reposant sur des données alternatives pour établir un profil de risque devraient eux aussi être obligés de communiquer les résultats de ces prêts aux agences d'évaluation du crédit.

Certains considèrent que les prêteurs alternatifs devraient être tenus de communiquer les données de crédit relatives aux prêts aux agences d'évaluation du crédit, que les prêts aient été remboursés (données de rapport positives) ou non (données de rapport négatives)[116]. Cette démarche pourrait permettre de définir des obligations réglementaires plus "équitables" pour des activités similaires (prêts), plutôt que d'appliquer des obligations différentes selon le type d'entité (une banque, par opposition à une entité non bancaire). Cela pourrait également élargir l'éventail d'informations disponibles sur les consommateurs, et ainsi combler les lacunes de l'écosystème des données et l'enrichir.

[116] Voir, par exemple: GPFI, *Use of Alternative Data*, à la note de bas de page 14.

Ces avantages potentiels doivent être évalués à la lumière de la manière dont le marché du crédit alternatif se développe. Les prêts effectués à l'aide de données alternatives et de décisions automatisées affichent souvent un faible montant (par exemple, pour permettre à quelqu'un de tenir jusqu'à la fin du mois). Leurs résultats peuvent donc avoir une utilité limitée. Le nouveau marché en pleine expansion des prêts automatisés, qui utilise des algorithmes propriétaires pour évaluer les emprunteurs sans antécédents de crédit traditionnels est également très novateur. Le fait d'exiger de ces nouveaux prêteurs qu'ils partagent leurs résultats en matière de prêts peut les priver de certains des avantages liés à leurs investissements et à leur rôle de pionniers. En outre, ces entreprises sont souvent des start-up entrepreneuriales qui peuvent être confrontées à de strictes obligations en matière d'information alors qu'elles cherchent à développer une activité risquée. Certaines ne s'appuient même pas sur les données des agences d'évaluation du crédit pour prendre des décisions en matière de prêts (elles s'appuient exclusivement sur des données alternatives), ce qui peut affaiblir la logique de réciprocité inhérente aux agences d'évaluation du crédit (selon laquelle ceux qui communiquent des données ont le droit de s'appuyer sur l'ensemble élargi de données agrégées communiquées par d'autres)[117].

Principes FEAT de l'Autorité monétaire de Singapour

Principe 3. Les données et les modèles utilisés pour prendre des décisions grâce aux technologies d'AIDA sont régulièrement examinés et validés à des fins d'exactitude et de pertinence [...].

Principe 4. Les décisions prises grâce aux technologies d'AIDA sont régulièrement réévaluées afin que les modèles se comportent comme prévu.

Version préliminaire des normes de la Smart Campaign relatives au crédit numérique

Indicateur 2.1.5.0
Les données et les analyses de souscription sont actualisées à chaque cycle de prêt afin de suivre l'évolution de la situation du client.

De ce fait, il est important d'examiner l'écosystème global des données du marché financier au fur et à mesure de son développement, tant en ce qui concerne l'exactitude des données utilisées dans les décisions automatisées que l'attribution des responsabilités quant à l'exactitude des informations dans les systèmes formels de communication des données de crédit et de manière plus générale.

Compte tenu du large éventail de données disponibles et de leurs sources et niveaux de fiabilité variables, de nombreux dilemmes politiques seront à résoudre quant au fonctionnement des lignes directrices relatives aux notions de clarté et de prévisibilité figurant dans le quatrième principe général relatif aux évaluations du crédit (le cadre juridique et réglementaire doit être suffisamment précis pour permettre aux prestataires de services, aux fournisseurs de données, aux utilisateurs et aux personnes concernées de prévoir les conséquences de leurs actes).

6.2 Protéger les usagers contre les biais et les traitements discriminatoires

Déductions et résultats décisionnels biaisés

Si l'une des préoccupations suscitées par les mégadonnées concerne la manière dont les données d'entrée (nom, âge, etc.) seront utilisées et protégées, les déductions qui résultent du traitement de ces données personnelles soulèvent également des inquiétudes. Les modalités et la précision des déductions que les

[117] Pour consulter un excellent article sur ces questions, voir: Blechman, J., "Mobile Credit in Kenya and Tanzania: Emerging Regulatory Challenges in Consumer Protection, Credit Reporting and Use of Customer Transactional Data". *African Journal of Information and Communication*, nº 17, novembre 2016. Disponible à l'adresse suivante: http://www.macmillankeck.pro/publications.html.

mégadonnées et l'apprentissage automatique permettront de tirer sur les individus et les groupes, ainsi que l'incidence de ces déductions sur les décisions, sont tout aussi importantes que l'exactitude des données d'entrée[118]. Certaines de ces déductions, qui permettent de prédire un comportement futur et sont difficiles à vérifier, peuvent déterminer la manière dont les individus sont perçus et évalués, et ainsi affecter leur vie privée, leur réputation et leur autodétermination.

Les lois sur la protection des données qui régissent la collecte, l'utilisation et le partage des données personnelles n'abordent généralement pas les résultats des modèles d'apprentissage automatique qui traitent ces données. L'une des préoccupations en matière de protection des données et de la vie privée concerne la manière de prévenir la discrimination. Le cinquième des Principes de haut niveau sur l'inclusion financière numérique stipule que les données ne doivent pas être utilisées de manière injuste et discriminatoire dans le cadre des services financiers numériques (par exemple, pour discriminer les femmes en matière d'accès au crédit ou aux assurances)[119].

De récents exemples de déductions impliquant les principales plates-formes Internet concernent l'orientation sexuelle, la santé physique et mentale, la grossesse, l'appartenance ethnique et les opinions politiques. Ces données peuvent être utilisées pour déterminer l'éligibilité d'une personne au crédit[120]. Le RGPD distingue des catégories particulières de données personnelles pour lesquelles les restrictions sont plus sévères. Si les données à caractère personnel se définissent comme "toute information se rapportant à une personne physique identifiée ou identifiable"[121], les "catégories particulières" de données à caractère personnel sont plus spécifiques. Elles concernent "l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique"[122].

Limitation du traitement des catégories particulières de données

La prise de décisions automatisée reposant sur des catégories particulières de données à caractère personnel n'est autorisée par le RGPD qu'avec le consentement explicite de la personne concernée ou si elle est "nécessaire pour des motifs d'intérêt public important, sur la base du droit de l'Union ou du droit d'un État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée"[123].

---

[118] Groupe de travail "Article 29" sur la protection des données, "Avis 03/2013 sur la limitation des finalités", 00569/13/FR WP 203, adopté le 2 avril 2013", 2013, p. 47. Disponible à l'adresse suivante: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf; Tene, O. et Polonetsky, J., "Big Data for All: Privacy and User Control in the Age of Analytics", *Northwestern Journal of Technology and Intellectual Property*, vol. 11, 2012, disponible à l'adresse suivante: https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip; Contrôleur européen de la protection des données, "Opinion 3/2018 on Online Manipulation and Personal Data", 19 mars 2018, p. 8-16, disponible à l'adresse suivante: https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf.

[119] G20, *High Level Principles of Digital Financial Inclusion*, p. 16. Disponible à l'adresse suivante: https://www.gpfi.org/publications/g20-high-level-principles-digital-financial-inclusion.

[120] Taylor, A. et Sadowski, J., "How Companies Turn Your Facebook Activity Into a Credit Score". *The Nation*, 15 juin 2015. Disponible à l'adresse suivante: https://www.thenation.com/article/how-companies-turn-your-facebook-activity-credit-score/.

[121] RGPD, article 4.

[122] RGPD, article 9, paragraphe 4.

[123] RGPD, article 22, paragraphe 4 et article 9, paragraphe 2, points a et g.

L'objectif de ces restrictions plus sévères concernant le traitement des catégories particulières de données est de fournir des moyens concrets de renforcer les autres lois interdisant la discrimination sur la base de ces informations, que ce soit dans la prestation de services publics ou privés, ou autrement. Le droit au respect de la vie privée vise à empêcher les divulgations susceptibles d'entraîner des discriminations et d'autres préjudices irréversibles[124].

Cependant, à l'ère des mégadonnées, des données non sensibles peuvent être utilisées pour déduire des données sensibles. Par exemple, le nom de la personne peut être utilisé pour déduire sa religion ou son lieu de naissance qui, à leur tour, peuvent être utilisés pour déduire son appartenance ethnique et d'autres données personnelles qui appartiennent aux catégories particulières de données. Les données d'achat peuvent révéler les antécédents d'achats de médicaments à partir desquels il est possible de déduire l'état de santé d'une personne, ce qui influerait par exemple sur les décisions relatives à son éligibilité à l'assurance maladie[125]. Les données démographiques et statistiques relatives à des groupes plus larges peuvent également être attribuées à certaines personnes. Il est donc possible que les données non sensibles méritent les mêmes protections que les données sensibles[126]. Le fait est que la distinction entre les données sensibles et non sensibles devient floue et d'une utilité discutable[127].

Il ne s'agit pas d'une simple question de définition. L'un des objectifs fondamentaux de la législation et de la réglementation en matière de protection des données et de la vie privée est de veiller à ce que les données ne soient pas utilisées à des fins de discrimination, en particulier à l'égard de groupes protégés qui en ont toujours été victimes. La nature même des mégadonnées et de l'apprentissage automatique compromet cet objectif. Comme l'ont récemment déclaré plusieurs spécialistes, la prise de décisions automatisée pourrait systématiser et dissimuler la discrimination, ce qui soulève de grandes inquiétudes[128].

Lorsque les algorithmes d'apprentissage automatique reposent sur des données d'entrée qui sont elles-mêmes tirées d'exemples passés, ils peuvent désavantager les groupes de population qui ont toujours été défavorisés. Ils peuvent donc refléter une discrimination passée, quelles qu'en aient été les raisons (par exemple, en raison de préjugés ou de biais implicites). Si ces décisions antérieures étaient elles-mêmes biaisées, les données servant aux processus d'apprentissage automatique peuvent perpétuer ou exacerber d'autres préjugés.

---

[124] Groupe de travail "Article 29" sur la protection des données, "Advice Paper on Special Categories of Data ("sensitive Data")", Ares(2011)444105-20/04/2011, p. 4. Disponible à l'adresse suivante: https://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf.

[125] Rouvroy, A., *Of Data and Men: Fundamental Rights and Freedoms in a World of Big Data.* Direction générale Droits de l'homme et État de droit du Conseil de l'Europe, 11 janvier 2016, p. 10. Disponible à l'adresse suivante: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020.

[126] S'agissant de la prise de décisions automatisée, le Groupe de travail "Article 29" a constaté que le profilage pouvait permettre d'obtenir des données sensibles par déduction à partir de données non sensibles qui, bien que ce n'était pas le cas au départ, appartiennent aux catégories particulières une fois combinées à d'autres données. Groupe de travail "Article 29" sur la protection des données, "Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679", note de bas de page 55, p. 15.

[127] Voir Zarsky, à la note de bas de page 16.

[128] Voir: Kroll, J., Huey, J., Barocas, S., Felten, E., Reidenberg, J., Robinson, D. et Yu, H., "Accountable Algorithms". *University of Pennsylvania Law Review*, 2017. Disponible à l'adresse suivante: https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9570&context=penn_law_review. Ohm, P. et Lehr, D., "Playing with the Data: What Legal Scholars Should Learn About Machine Learning". *University of California Davis Law Review*, 2017. Disponible à l'adresse suivante: https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

La solvabilité d'une personne peut être évaluée en fonction non seulement de ses attributs, mais aussi de ceux de son réseau social. En 2015, Facebook a déposé un brevet qui permettait, entre autres, de filtrer les demandes de prêt selon la cote de crédit moyenne des amis d'un demandeur de prêt, par rapport au score de crédit minimal prescrit[129]. Cette situation peut entraîner un risque de discrimination, voire d'exclusion financière, si les amis d'un demandeur appartiennent majoritairement à un groupe de population à faible revenu, même si les caractéristiques propres au demandeur le qualifiaient pour le prêt[130]. Le risque est qu'en s'appuyant sur des données antérieures, ces technologies facilitent l'accès des populations les plus riches aux services financiers et entravent au contraire l'accès des groupes minoritaires, qui n'y avaient déjà pas accès par le passé, ce qui "automatise les inégalités"[131].

La discrimination peut également être intégrée aux modèles d'apprentissage automatique dans la "sélection des paramètres", c'est-à-dire les choix effectués lors de leur conception concernant les données à prendre en compte. Si un modèle peut ne pas tenir explicitement compte de l'appartenance à une catégorie protégée (par exemple, le genre, l'appartenance ou l'origine ethnique, la religion), en particulier si cela est illégal, il peut néanmoins s'appuyer sur des données qui permettent de faire des déductions sur l'appartenance à ladite catégorie protégée. Les codes postaux sont un exemple couramment utilisé, car certaines zones abritent un pourcentage élevé de la population provenant d'un groupe ethnique particulier.

Un autre problème se pose lorsque le modèle d'apprentissage automatique ne prend pas en compte un ensemble de facteurs suffisamment large pour garantir que les membres d'un groupe protégé sont évalués avec autant de précision que les autres. Un modèle peut disposer d'un nombre limité de données de crédit sur les membres d'un groupe défavorisé parce qu'ils sont moins nombreux à avoir emprunté par le passé. Si les algorithmes sont entraînés à partir d'un plus grand volume de données d'entrée provenant d'un certain groupe (plutôt que d'un autre), ils peuvent produire des résultats favorisant ce groupe au détriment d'un autre.

En outre, les modèles d'apprentissage automatique pourraient être utilisés pour masquer délibérément la discrimination. Cela pourrait se produire si les données d'apprentissage sont volontairement faussées ou si des substituts d'une catégorie protégée sont utilisés exprès pour produire des résultats discriminatoires.

Les techniques permettant d'éliminer les biais fondés sur des attributs protégés visent à garantir que l'étiquette que l'on colle à un individu ne tient pas compte de ces attributs[132]. Cependant, même si lesdits attributs ne sont pas explicitement pris en compte, des attributs corrélés (indirects) peuvent être intégrés à l'ensemble de données, ce qui donne des résultats potentiellement discriminatoires. Il est difficile de tenir compte de cet aspect dans le cadre de l'apprentissage automatique, mais des tests ont été mis au point pour évaluer l'incidence d'une décision automatisée sur différents groupes protégés[133].

---

[129] http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=9100400.PN.&OS=PN/9100400&RS=PN/9100400.

[130] Zim, J., "The Use of Social Data Raises Issues for Consumer Lending". *Miami Business Law Review*, 2016. Disponible à l'adresse suivante: https://business-law-review.law.miami.edu/social-data-raises-issues-consumer-lending/.

[131] Eubanks, V., *Automating Inequality.* St Martin's Press, 2018.

[132] Hardt, M., Price, E. et Srebro, N., *Equality of opportunity in supervised learning,* Advances in Neural Information Processing Systems, 2017; Chouldechova, A., *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments,* CoRR, 2017.

[133] L'impact disparate a été défini à l'aide de la "règle des 80%", de sorte que, lorsqu'un ensemble de données possède un attribut protégé X (par exemple, l'appartenance ethnique, le genre, la religion, etc.) et un résultat binaire à prédire C (par exemple, "embauchera"), l'ensemble de données a un impact disparate si:

Dans certains pays, un biais non intentionnel peut tout à fait être illégal s'il a un "impact disparate", notamment lorsque les résultats d'un processus de sélection sont très différents pour une catégorie de personnes protégées (par exemple, en fonction du genre, de l'appartenance ou de l'origine ethnique, ou de la religion), par rapport à d'autres groupes, alors que le processus semble quant à lui être neutre. La notion d'impact disparate a été développée à partir d'une décision de la Cour suprême des États-Unis, en 1971[134], qui a constaté que certains résultats de tests d'intelligence et certains diplômes d'études secondaires dépendaient largement de l'appartenance ethnique, ce qui peut conduire à des décisions d'embauche discriminatoires[135]. La Cour suprême des États-Unis a récemment réaffirmé cette théorie juridique lorsqu'en 2015, elle a stipulé qu'un plaignant pouvait établir un commencement de preuve pour discrimination en vertu de la loi sur le logement équitable (*Fair Housing Act*), sans avoir à prouver que ladite discrimination était intentionnelle s'il apportait la preuve statistique qu'une politique gouvernementale provoquait un impact disparate[136].

Il est plus difficile d'évaluer l'impact disparate, et partant les biais, lorsque des ordinateurs sont impliqués. Il peut par ailleurs s'avérer difficile, voire impossible, de divulguer et d'expliquer le processus de sélection par algorithme. Cependant, lorsqu'il peut être démontré qu'un modèle produit des résultats discriminatoires, il est possible qu'il enfreigne également les lois interdisant la discrimination, bien que cela puisse être difficile à prouver, notamment lorsque des justifications telles que la nécessité commerciale entrent en jeu[137].

On peut parler de sélection discriminatoire sans que des groupes protégés soient impliqués. Par exemple, lorsque les algorithmes des services financiers numériques déduisent à partir des données utilisateur qu'une personne a des problèmes de liquidités financières, les prêteurs sur salaire peuvent être en mesure de cibler les personnes vulnérables avec des publicités et des offres de prêts à des taux d'intérêt et des frais élevés. La concurrence d'entreprises comme ZestFinance peut effectivement faire baisser le coût des prêts octroyés à ces groupes, mais si une sélection discriminatoire entraîne des résultats négatifs pour un individu, cela pourrait soulever des inquiétudes[138].

$$\frac{\Pr(C = YES|X = 0)}{\Pr(C = YES|X = 1)} \leq \tau = 0.8$$

pour la catégorie OUI à résultat positif et l'attribut protégé majoritaire 1 où Pr(C = cjX = x) désigne la probabilité conditionnelle (évaluée sur D) que le résultat de la catégorie soit c 2 C compte tenu de l'attribut protégé x 2 X. Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C. et Venkatasubramanian, S., *Certifying and removing disparate impact*, dans KDD, 2015. Disponible à l'adresse suivante: http://sorelle.friedler.net/papers/kdd_disparate_impact.pdf.

[134] Cour suprême des États-Unis, *Griggs c. Duke Power Co.*, 401 U.S. 424, 8 mars 1971.

[135] La Cour suprême des États-Unis a estimé que la décision d'embauche de Duke Power était illégale si elle entraînait un "impact disparate" en fonction de l'appartenance ethnique, même si ce facteur n'avait pas explicitement influencé ladite décision. Elle a ainsi empêché Duke Power d'utiliser les résultats des tests d'intelligence et les diplômes d'études secondaires, des qualifications hautement corrélées à l'appartenance ethnique, pour prendre des décisions d'embauche. La doctrine juridique de l'impact disparate, qui a été développée à partir de cet arrêt, est la principale théorie juridique utilisée pour déterminer si un acte de discrimination est volontaire ou non aux États-Unis. Duke Power n'a pas pu prouver que les tests d'intelligence ou les diplômes exigés étaient pertinents dans le cadre de ses offres d'emploi.

[136] *Texas Dep't of Housing and Community Affairs c. Inclusive Communities Project*, 135 S. Ct. 2507, 2015.

[137] Barocas, S. et Selbst, A. D., "Big Data's Disparate Impact". *California Law Review*, vol. 104, 2016, p. 671. Disponible à l'adresse suivante: http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf.

[138] Lohr, S., "Big Data Underwriting for Payday Loans". *The New York Times*, 19 janvier 2015. Disponible à l'adresse suivante: https://bits.blogs.nytimes.com/2015/01/19/big-data-underwriting-for-payday-loans/.

La lutte contre les tendances à la discrimination

L'une des manières d'aborder la tendance potentielle de l'apprentissage automatique à la discrimination consiste à prendre en compte le caractère aléatoire des

Principes FEAT de l'Autorité monétaire de Singapour

1. Les individus ou groupes d'individus ne sont pas systématiquement désavantagés par les décisions prises grâce aux technologies d'AIDA, sauf si ces décisions peuvent être justifiées.

3. Les données et les modèles utilisés pour prendre des décisions grâce aux technologies d'AIDA sont régulièrement examinés et validés pour limiter les biais involontaires.

Version préliminaire des normes de la Smart Campaign relatives au crédit numérique

Indicateur 5.2.1.0

En fonction de son appartenance ethnique, son genre, son âge, son handicap, son affiliation politique, son orientation sexuelle, sa caste et sa religion, une personne peut être placée dans une catégorie protégée.

Indicateur 5.2.3.0

Les algorithmes sont conçus pour réduire le risque de discrimination des consommateurs liée aux catégories protégées.

Indicateur 5.2.3.1

Après une phase d'apprentissage initiale, le fournisseur effectue une analyse des liens entre les variables non discriminatoires et les variables discriminatoires afin de vérifier l'absence de biais involontaire dans les décisions de crédit automatisées.

Indicateur 5.2.3.2

Si le fournisseur confie le développement de l'algorithme à une tierce partie, il doit exiger de cette dernière qu'elle respecte les normes exposées à l'indicateur ci-dessus. La tierce partie communique au fournisseur les informations suivantes: paramètres et documentation de l'algorithme, supports de formation fournis à l'équipe et documents relatifs à l'historique des tests antérieurs (comprenant notamment la date, une description et le résultat de chaque test, les éléments de discrimination identifiés et les mesures correctives prises, le cas échéant).

données[139]. Par exemple, un algorithme d'apprentissage automatique employé pour l'octroi de crédits peut être entraîné grâce à des données initiales qui indiquent qu'un certain groupe (défini en fonction, par exemple, du code postal, du genre ou de l'appartenance ethnique des personnes) a tendance à rassembler des débiteurs moins fiables. Si le modèle permettait d'octroyer un crédit à d'autres groupes, il pourrait en résulter une prophétie autoréalisatrice selon laquelle les caractéristiques des personnes qui règlent leurs dettes sont liées à leur non-appartenance au groupe protégé. L'incorporation d'un élément aléatoire dans le modèle, de sorte que certains individus qui, d'après les déductions tirées, ne seraient généralement pas considérés comme des débiteurs fiables bénéficient néanmoins d'un crédit, pourrait permettre au modèle de vérifier la validité des hypothèses initiales. L'introduction de données évolutives reflétant mieux le monde réel peut contribuer à améliorer l'équité et la précision globales du système.

Une autre approche proposée consiste à sélectionner ou à modifier les données d'entrée de manière à ce que le produit réponde aux critères d'équité appliqués par le système. Pour ce qui est des groupes minoritaires, des échantillons d'apprentissage supplémentaires peuvent être sélectionnés afin d'éviter de trop mettre en avant leur statut de minorité. D'autres méthodes peuvent être adoptées pour garantir la parité statistique entre les groupes[140]; l'important est de veiller à ce qu'elles soient incorporées au modèle à l'étape de conception, même en utilisant l'intelligence artificielle pour surveiller l'utilisation de l'intelligence artificielle.

Dans certains cas, on pourrait s'attendre à ce qu'il y ait une incitation commerciale à éliminer les biais. La partialité ne nuit pas seulement à la réputation d'un service, elle peut instaurer une économie commerciale sous-optimale pour le fournisseur de services. Si le code postal d'un demandeur de prêt affecte son score et entraîne le rejet de sa demande de prêt alors qu'il dispose d'un revenu sain, d'un faible niveau d'endettement et d'autres attributs positifs, le prêteur a manqué une occasion d'accorder un prêt rentable.

Sur un marché statique idéal où les fournisseurs sont en concurrence pour le même service et sont à même de l'améliorer pour accroître leur part de marché, on pourrait s'attendre à ce que les développeurs améliorent les algorithmes au fil du temps pour éliminer les biais. Toutefois, sur un marché dynamique où de nouveaux modèles et services sont constamment mis au point et de nouvelles données sont ajoutées en permanence, il se peut que les biais soient corrigés, mais qu'un modèle soit mis à jour ou remplacé par un nouveau qui reflétera peut-être un autre biais, ce qui répète le problème. Les entreprises peuvent également se concentrer davantage sur une croissance rapide pour conquérir le nouveau marché, tout en considérant l'impact discriminatoire sur les groupes protégés comme un aspect de moindre importance. Même si l'on peut s'attendre à ce que le marché affine avec le temps les algorithmes afin de réduire les biais, dans de nombreux cas, il est tout simplement inacceptable, socialement et politiquement parlant, d'autoriser les biais liés à l'appartenance ou à l'origine ethnique, ou encore au genre.

Une question clé est de savoir dans quelle mesure l'industrie doit assumer la responsabilité et le coût liés à l'identification des biais, en utilisant les données à sa disposition pour repérer toute discrimination

---

[139] Voir: "Accountable Algorithms", à la note de bas de page 128. Ohm, P. et Lehr, D., "Playing with the Data: What Legal Scholars Should Learn About Machine Learning". *University of California Davis Law Review*, 2017. Disponible à l'adresse suivante: https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

[140] Voir: "Accountable Algorithms", à la note de bas de page 128. Ohm, P. et Lehr, D., "Playing with the Data: What Legal Scholars Should Learn About Machine Learning". *University of California Davis Law Review*, 2017. Disponible à l'adresse suivante: https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

éventuelle. Lorsque la prise de décisions automatisée est à l'origine de discriminations et de préjudices illicites en vertu des lois en vigueur, les entreprises qui recourent à un tel processus peuvent utiliser des outils (et, selon certaines lois, en avoir l'obligation), afin de s'assurer que l'utilisation des données n'exacerbera pas les préjugés historiques, ainsi que des méthodes de traitement des données qui évitent d'utiliser des variables de substitution pour les catégories protégées. En outre, il peut s'avérer nécessaire de faire examiner les résultats des algorithmes par des êtres humains. Il est également possible d'utiliser les données disponibles pour repérer les discriminations, et d'obliger les entreprises à le faire par voie réglementaire.

Même si les résultats n'enfreignent pas les lois interdisant la discrimination fondée sur l'appartenance ethnique, la religion ou tout autre facteur propre à une catégorie protégée, le préjudice injuste causé aux personnes peut amener les entreprises à adopter des cadres déontologiques et de "bonnes pratiques" pour ajuster les algorithmes, afin de veiller au contrôle et à l'évaluation des résultats. Certaines mesures d'atténuation peuvent consister à donner aux personnes la possibilité (ou le droit) de recevoir une explication concernant les décisions automatisées (voir section 7.2) et à recourir à une évaluation de l'impact des mesures de protection des données (voir section 8).

D'autres approches ont été suggérées, notamment l'évaluation aléatoire, de temps à autre, par les organismes de protection des consommateurs, des systèmes de notation des fournisseurs de services financiers (et des prestataires de soins de santé, des établissements d'enseignement et d'autres organismes qui prennent régulièrement des décisions concernant les individus). Des scénarios hypothétiques pourraient être utilisés afin d'évaluer si les modèles utilisent effectivement des statistiques de substitution pour les groupes protégés (appartenance ethnique, genre, religion, handicap, etc.). Un audit de ce type pourrait encourager les entreprises à se prémunir contre ces risques[141].

Tarification différenciée et autres conditions

Grâce aux données disponibles, un fournisseur de services financiers est en mesure de mieux évaluer le risque que représente un consommateur, et donc de lui proposer des services qui ne lui seraient peut-être pas accessibles autrement. La disponibilité d'un vaste éventail de données sur un consommateur entraîne toutefois une asymétrie d'information, le fournisseur en sachant plus sur le consommateur que le consommateur sur le fournisseur. Le fournisseur peut profiter de cette situation et appliquer ce que les économistes appellent une "tarification différenciée", une pratique qui revient à facturer, pour le même produit, des prix différents à différents consommateurs.

La tarification différenciée est toutefois une pratique courante qui présente souvent des avantages pour les consommateurs. Prenons l'exemple des billets de train, qui sont souvent vendus à un prix réduit aux étudiants et aux retraités. Cependant, elle peut également susciter un sentiment d'injustice lorsque certains groupes de population sont amenés à payer des prix plus élevés en raison de leur profil (situation géographique, etc.)[142].

Dans le cadre des services financiers, la tarification différenciée repose principalement sur le profil de risque du consommateur. Une tarification basée sur le risque peut améliorer l'efficacité économique en

---

[141] Citron, D., "Technological Due Process". *Washington University Law Review*, vol. 85, 2007, p. 1249-1313. Disponible à l'adresse suivante: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1012360.

[142] Angwin, J. et Larson, J., "The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review". *ProPublica*, 1er septembre 2015.

décourageant les comportements à risque, en récompensant les personnes qui n'ont pas d'antécédents d'activités illicites ou d'accidents de la route. Elle peut améliorer l'accès à l'assurance en limitant l'antisélection, une pratique selon laquelle seules les personnes présentant un profil à haut risque peuvent souscrire à un prix uniforme. La tarification différenciée des produits d'assurance peut toutefois entraîner des injustices lorsque les facteurs de risque échappent au contrôle d'un individu, comme c'est le cas pour l'assurance maladie.

Les mégadonnées peuvent favoriser la tarification différenciée, car elles permettent de tirer des déductions à partir des données personnelles, notamment sur la mesure dans laquelle un individu a besoin de tel ou tel service, sa capacité à payer et sa sensibilité au prix. L'ordinateur peut estimer un prix aussi proche que possible du montant maximal que le consommateur profilé serait prêt à payer. En raison d'une asymétrie d'information, le consommateur n'en sait pas assez sur le fournisseur pour négocier le prix jusqu'au montant minimal que le fournisseur serait prêt à accepter (par exemple, pour obtenir un retour sur investissement raisonnable).

Sur un marché dynamique, on s'attendrait à ce que la concurrence exerce une pression à la baisse sur le prix du fournisseur, diminuant la marge de profit. Toutefois, lorsque la tarification différenciée désavantage des personnes déjà défavorisées, cela soulève des problèmes d'ordre politique. Une personne peut avoir désespérément besoin d'un service financier, et donc être prête à payer un prix plus élevé que la normale. Un prêteur peut être en mesure de facturer un prix plus élevé qu'en temps normal, qui ne reflète pas tant le risque supérieur de défaut de remboursement que l'urgence de l'emprunteur. Cela peut porter préjudice aux personnes et aux familles à faible revenu.

La tarification différenciée peut également devenir discriminatoire lorsque les prix sont fixés en fonction de critères qui, bien qu'apparemment objectifs, entraînent un traitement défavorable à l'égard des groupes protégés. Par exemple, si un algorithme fixe des prix plus élevés pour les consommateurs dont le code postal provient d'un quartier qui, par le passé, a connu des niveaux de défaut de remboursement plus élevés que ceux d'autres quartiers, les personnes dont les autres attributs n'indiquent pas un plus grand risque peuvent néanmoins faire face à des prix plus élevés que la normale.

Certains groupes de population qui ont toujours été défavorisés ont des attributs en commun (tels que le code postal). Les personnes possédant ces attributs peuvent donc être victimes de discrimination, même si ceux-ci n'influent pas sur leur solvabilité. Par exemple, une personne ayant un bon salaire et peu de dettes peut être traitée de manière défavorable parce qu'elle vit dans une communauté qui a connu par le passé des ratios dette/revenu supérieurs à la moyenne (il peut aussi bien s'agir de ses amis sur les médias sociaux, du fait qu'elle a le même médecin qu'eux ou qu'elle fasse ses courses dans des magasins à prix réduit). Les modèles d'apprentissage automatique contribuent donc aux tendances à l'automatisation des processus économiques susceptibles d'exacerber les inégalités au fil du temps[143].

6.3 Protéger les usagers en cas de violation des données et de réidentification

Les grands volumes de données détenues et transférées par les acteurs du domaine des mégadonnées risquent de compromettre la sécurité des données, et donc d'entraîner des risques pour la vie privée des

[143] Harris, K., Kimson, A. et Schwedel, A., "Labor 2030: The Collision of Demographics, Automation and Inequality", *Bain & Company Report*, 7 février 2018. Disponible à l'adresse suivante: http://www.bain.com/publications/articles/labor-2030-the-collision-of-demographics-automation-and-inequality.aspx.

usagers. Même lorsque la quantité de données détenues sur un individu est réduite au minimum, son identité peut être découverte par rétro-ingénierie, à partir d'un nombre même restreint de points de données, ce qui risque de porter atteinte à sa vie privée[144]. Ce risque existe lorsque les données peuvent être obtenues par des tiers, qui y auront eu accès bien que n'y étant pas autorisés (violation des données) ou à qui l'on aura communiqué des informations avec l'accord de l'entreprise qui contrôle ou traite les données. Dans les deux cas, les mesures visant à protéger la divulgation de données relatives à des personnes identifiables comprennent des processus de désidentification, de pseudonymisation et d'anonymisation. Ces mesures et les difficultés qui entravent leur application dans le contexte des mégadonnées sont examinées dans la présente section. La section 1.1 aborde quant à elle le rôle et la réglementation des intermédiaires indépendants qui acquièrent des données sur contrat sur le marché des données.

Les limites des processus de désidentification, de pseudonymisation et d'anonymisation

La vie privée peut être protégée à des degrés divers par l'utilisation de technologies d'amélioration de la confidentialité[145], telles que la désidentification, qui consiste à supprimer ou à ajouter du bruit aux informations d'identification directe et indirecte dans un ensemble de données, ou à introduire d'autres obstacles (rendant l'identification d'une personne statistiquement improbable)[146]:

- Les données d'identification *directe* permettent d'identifier une personne sans informations supplémentaires ou grâce à des informations du domaine public (par exemple, son nom, son numéro de téléphone, son adresse électronique, sa photographie, son numéro de sécurité sociale ou ses identifiants biométriques).

- Les données d'identification *indirecte* comprennent les attributs qui peuvent être utilisés pour identifier une personne (son âge, sa localisation et autres caractéristiques personnelles uniques).

Alors que la désidentification implique la suppression de ces deux catégories de données, la pseudonymisation ne supprime que les données d'identification directe, de sorte que les données à caractère personnel ne peuvent être attribuées à une personne spécifique sans l'utilisation d'informations complémentaires. Ces informations supplémentaires sont conservées séparément et protégées par des mesures techniques et administratives visant à empêcher une telle attribution[147]. Le processus de pseudonymisation de base n'est pas compliqué; il suffit de substituer des attributs alternatifs:

[144] Voir, par exemple: Ohm, P., "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization". *UCLA Law Review*, vol. 157, 2010, p. 1701, 1716-1727.
[145] Future of Privacy Forum, "A Visual Guide To Practical Data De-Identification".
[146] Voir: Cavoukian, A. et El-Emam, K., *De-Identification Protocols: Essential for Protecting Privacy*, Commissaire à l'information et à la protection de la vie privée de l'Ontario, 2014; et Commissaire à l'information et à la protection de la vie privée de l'Ontario, "De-Identification Centre", accessible à l'adresse suivante https://www.ipc.on.ca/wp-content/uploads/resources/pbd-de-identifcation_essential.pdf .
[147] RGPD, article 4, paragraphe 5.

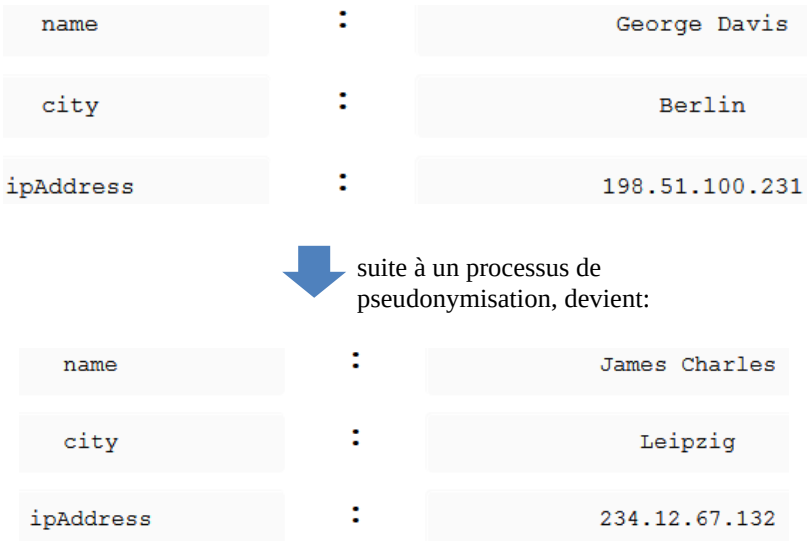

Figure 3 – Processus de pseudonymisation. Source: KIProtect

La désidentification est l'un des moyens par lesquels les organisations peuvent se conformer aux exigences de "minimisation des données" des lois sur la protection des données, c'est-à-dire ne recueillir, ne conserver et n'utiliser que les données personnelles strictement nécessaires et pertinentes pour l'objectif défini (voir section 5.1).

La désidentification élimine rarement le risque de réidentification. Une réidentification est possible lorsque le processus de désidentification a été mis en œuvre ou contrôlé de manière inadéquate, ou qu'il est possible de relier les données désidentifiées à des données personnelles déjà connues ou à des informations accessibles au public. Une désidentification efficace nécessite une excellente compréhension des données et de l'écosystème des données au sens large, y compris des raisons pour lesquelles des parties adverses pourraient chercher à réidentifier certaines personnes, et des moyens qu'elles pourraient employer.

Certains experts critiquent l'inefficacité de la désidentification et la promotion d'un faux sentiment de sécurité reposant sur des modèles irréalistes et artificiellement limités de ce que pourrait faire une partie adverse[148]. Reprenons un célèbre exemple de 1997: en reliant des données de santé dépourvues d'identifiants personnels à des données d'inscription sur les listes électorales accessibles au public, il a été possible d'identifier le gouverneur William Weld du Massachusetts et de faire ainsi le rapprochement avec son dossier médical. (Celui-ci avait auparavant assuré à ses électeurs que leurs données de santé étaient parfaitement confidentielles[149].)

[148] Narayanan, A. et Felten, E. W. (Princeton), *No silver bullet: De-identification still doesn't work*, 2014. Disponible à l'adresse suivante: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf.

[149] Ohm, P., "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization". *UCLA Law Review*, vol. 57, 2010, p. 1701. Disponible à l'adresse suivante: https://www.uclalawreview.org/pdf/57-6-3.pdf.

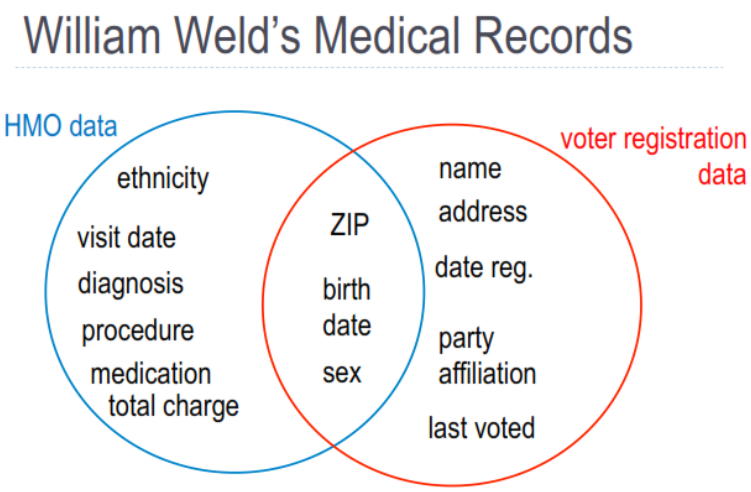

Figure 2 – Identification du gouverneur Weld à partir de quatre attributs

Une étude réalisée en 2013 révèle que 95% des traces de mobilité sont individualisables grâce à quatre points spatiotemporels aléatoires (données et heure), et que plus de 50% des utilisateurs sont individualisables à partir de deux points choisis au hasard (qui sont généralement le domicile et le lieu de travail) [150] . Des études ultérieures ont obtenu des résultats similaires en utilisant de grands ensembles de données (par exemple, un million de personnes en Amérique latine) et en appliquant cette méthodologie aux données de transactions bancaires. Elles ont montré que quatre points spatiotemporels suffisaient pour identifier 90% des utilisateurs de cartes de crédit[151].

Des données plus riches permettent d'identifier un individu grâce à un ensemble de champs ou d'attributs, par exemple le code postal, la date de naissance et le genre.

Les données de géolocalisation comportent des risques particuliers d'identification ou de réidentification des personnes. Il est possible de combiner les données d'un utilisateur liées à un identifiant permanent et non unique avec d'autres données afin de dresser le profil détaillé d'une personne. Même les données de géolocalisation peuvent, à elles seules, être utilisées pour identifier une personne, car les deux endroits où elle se trouve le plus souvent sont généralement son domicile et son lieu de travail. Les données sensibles concernant une personne, qui renseignent par exemple sur une condition médicale particulière, peuvent être identifiées à l'aide des lieux fréquentés (clinique d'avortement, mosquée, etc.).

Il existe des mesures pour réduire ces risques, par exemple: accepter des aperçus plutôt que des ensembles de données complets; n'accepter que des données qui ont déjà été agrégées ou désidentifiées; appliquer des filtres supplémentaires lorsque les données proviennent de certains dispositifs; n'accepter que des données géoclôturées; supprimer le domicile, le lieu de travail et les lieux sensibles; limiter la durée de conservation des données; et "brouiller" ou "flouter" les ensembles de données.

L'anonymisation implique l'élimination ou la transformation des données d'identification directe et indirecte. Alors que la pseudonymisation et la désidentification impliquent des procédures et des contrôles techniques, institutionnels et juridiques pour empêcher les employés et les parties tierces (tels que les chercheurs) de réidentifier les personnes, l'anonymisation – une fois réalisée – ne nécessite

[150] De Montjoye, Y.-A., *et al.*, "Unique in the Crowd: The privacy bounds of human mobility". *Scientific Reports*, vol. 3, 2013.
[151] Initiative Global Pulse des Nations Unies, *Mapping the risk-utility landscape of mobile phone data for sustainable development & humanitarian action*, 2015; Song, Y., Dahlmeier, D. et Bressan, S., *Not so unique in the crowd: a simple and effective algorithm for anonymizing location data*, ACM PIR, 2014; De Montjoye, Y.-A., Radaelli, L. et Singh, V. K., "Unique in the shopping mall: On the reidentifiability of credit card metadata", *Science*, vol. 347, nº 6221, 2015, p. 536-539.

aucune de ces mesures supplémentaires. Elle limite cependant l'utilité des données. Plus les données sont riches, plus elles sont utiles.

Amélioration des approches en matière de risque de réidentification

On assiste à l'émergence de technologies et de critères qui visent à préserver la richesse des données tout en réduisant l'identifiabilité des individus. Par exemple, la "confidentialité différentielle" a gagné en popularité depuis qu'Apple a annoncé s'en servir pour anonymiser les données de ses utilisateurs[152]. La confidentialité différentielle permet d'évaluer la qualité de l'anonymisation des données. Elle permet de quantifier le volume d'informations que la méthode d'anonymisation fera fuir au sujet d'un individu donné ajouté à un ensemble de données à l'aide de cette méthode. Elle repose sur des compromis entre utilité et commodité, en introduisant un bruit aléatoire pour éliminer les différences en matière de divulgation entre un individu dont les données sont incluses dans l'analyse des mégadonnées et celui qui choisit de s'en retirer[153].

Lorsque le nombre d'individus concernés est suffisamment élevé, alors que le bruit statistique légèrement biaisé masque les données des individus, le bruit s'étale sur un grand nombre de points de données, ce qui permet de détecter des formes et de dégager des informations significatives. Cela offre un moyen d'évaluer les préjudices cumulatifs sur plusieurs utilisations, ce qui permet de meilleures discussions et décisions au sujet des compromis entre la vie privée et l'utilité statistique.

Des mécanismes de confidentialité différentielle des bases de données peuvent rendre les données confidentielles largement disponibles aux fins d'une analyse précise, sans avoir recours à des salles blanches, à des accords d'utilisation ou des plans de protection des données, ni à des aperçus limités. Cette approche résout ainsi le paradoxe consistant à ne rien apprendre sur un individu tout en obtenant des informations utiles sur toute une population[154].

Le contrôle de la divulgation statistique et des déductions, l'exploration de données respectueuse de la vie privée et l'analyse de données privées sont d'autres techniques algorithmiques qui peuvent être appliquées à de grandes bases de données en utilisant des méthodes statistiques en vue de gérer la confidentialité.

Le marché des services de désidentification, de pseudonymisation et d'anonymisation est en pleine expansion. Par exemple, la société allemande KIProtect[155] permet aux entreprises qui travaillent avec de grands ensembles de données de protéger ces dernières, en intégrant par le biais d'API le traitement des données de l'entreprise cliente pour repérer et protéger les données privées ou sensibles en les transformant à l'aide de techniques de pseudonymisation, d'anonymisation et de chiffrement. La prise en

---

[152] Apple, "Differential Privacy". Disponible à l'adresse suivante: https://images.apple.com/privacy/docs/Differential_Privacy_Overview.pdf.

[153] Il s'agit d'une solide garantie de confidentialité pour les entrées individuelles dans une fonction (aléatoire) ou une séquence de fonctions, que nous appellerons mécanisme de confidentialité. De manière informelle, avec cette garantie, le comportement du mécanisme demeure essentiellement inchangé, qu'une personne choisisse d'intégrer l'ensemble de données ou de refuser d'en faire partie. Prévue pour l'analyse statistique des données de santé ou de recensement, par exemple, cette définition protège la vie privée des individus, et de petits groupes d'individus, tout en permettant des résultats très différents avec des ensembles de données très différents. Dwork, C., "The differential privacy frontier", dans *Theory of Cryptography Conference.* Springer, Berlin, Lecture Notes in Computer Science vol. 5444, 2009, p. 496-502.

[154] Dwork, C., *Differential Privacy.* Compte rendu du 33e Colloque international sur les automates, les langages et la programmation, 2006.

[155] Voir www.kiprotect.com.

charge de nombreux types de données et de technologies de stockage (par exemple, Apache Kafka et Google Firebase) permet une utilisation dans un large éventail de contextes. La disponibilité croissante de ces fournisseurs de services signifie que les entreprises traitant des données peuvent externaliser une grande partie de leurs besoins en matière de confidentialité, réduisant ainsi la charge que représente la mise en place de leur propre capacité interne en matière de protection de la vie privée, qui n'est pas leur activité principale.

Les méthodes de désidentification, de pseudonymisation et d'anonymisation doivent non seulement être incluses dans le codage de la gestion des ensembles de données, mais aussi dans l'organisation administrative. Ainsi, Apple soumet les données de l'utilisateur à un processus de confidentialité différentielle sur l'appareil de ce dernier avant de les anonymiser (en supprimant les adresses IP et autres métadonnées), ainsi que de les recueillir, de les agréger et de les analyser. Les étapes d'ingestion et d'agrégation sont toutes deux réalisées dans un environnement à accès limité, de sorte que même les données privatisées ne sont pas facilement accessibles aux employés d'Apple[156].

Outre ce type de mesures, une politique de "séparation des tâches" peut réduire les risques pour la vie privée lors du traitement des données personnelles. Cela restreint le pouvoir de chaque administrateur à un rôle donné, les fonctions d'autres administrateurs étant limitées de la même manière, ce qui réduit le risque posé par un administrateur malveillant. Dans le même ordre d'idées, une politique appliquant le principe de "moindre privilège" viserait à garantir que chaque administrateur ne dispose que des pouvoirs nécessaires aux fonctions qui lui ont été déléguées.

En fin de compte, la difficulté d'empêcher la réidentification montre qu'il n'est peut-être pas judicieux d'avoir une vision manichéenne de la désidentification, et que le débat sur l'efficacité de ces techniques doit être abordé d'une manière plus nuancée, en acceptant l'idée que la désidentification peut parfois fournir des réponses acceptables[157]. En effet, Cynthia Dwork suggère que l'utilisation continue de données exactes finira par porter atteinte à la vie privée et que les techniques susvisées atténuent le risque sans toutefois l'éliminer[158]:

> *L'utilité des données finira par s'éroder: la loi fondamentale de la récupération de l'information dispose que des réponses trop précises à un trop grand nombre de questions anéantiront la confidentialité de manière spectaculaire. L'objectif de la recherche algorithmique sur la confidentialité différentielle est de repousser cette fatalité aussi longtemps que possible.*

Dans cette optique, la réglementation pourrait chercher à s'appuyer moins sur le fait d'informer les usagers que leurs données seront recueillies, analysées et partagées, ainsi que sur l'obtention de leur consentement à cet égard, et davantage sur la garantie que les technologies d'amélioration de la confidentialité seront toujours intégrées dans le traitement des mégadonnées et des données de l'apprentissage automatique, et mises à jour pour faire face aux nouveaux défis. Pour y parvenir, il pourra s'avérer nécessaire de mettre en place des mesures incitatives dans le cadre de la législation qui impliquent une responsabilité en cas de violation des données, faisant essentiellement peser le fardeau

---

[156] "Differential Privacy", à la note de bas de page 152.
[157] Professeur Casanovas, P., De Koker, L., Mendelson, D. et Professeur Watts, D., "Regulation of Big Data: Perspectives on Strategy, Policy, Law and Privacy". *Health and Technology*, 2017. Disponible à l'adresse suivante: https://ssrn.com/abstract=2989689.
[158] "Differential Privacy", note de bas de page 154.

économique non pas sur les usagers en obtenant leur consentement, mais sur les organisations qui recueillent, utilisent et partagent les données.

6.4 Protéger les usagers contre la circulation de données à caractère personnel les concernant

Les mégadonnées et l'apprentissage automatique sont rendus possibles non seulement par l'offre de données issues des activités en ligne et la demande des fournisseurs de services qui en dépendent, mais aussi par des intermédiaires – les courtiers en données indépendants qui font le commerce des données personnelles. Il en résulte un nombre considérable de sources de données, ainsi que de méthodes de collecte et de formats.

La communication des données à caractère personnel comporte plusieurs risques pour les usagers. Le transfert de données d'une entité à une autre augmente le risque de violation en raison de la hausse du nombre de parties en leur possession, ainsi que des vulnérabilités du processus de transfert en lui-même. Des données sensibles et confidentielles peuvent être obtenues par des parties tierces qui n'en ont pas l'autorisation, entraînant un risque d'usurpation d'identité, de marketing intrusif et d'autres atteintes à la vie privée.

Le transfert même des données à une partie tierce peut être une chose à laquelle l'usager ne s'attendait pas lorsqu'il a initialement communiqué ses données à une entreprise, par exemple en accédant à un service ou en naviguant simplement sur Internet. Enfin, comme indiqué à la section 4.4, la prolifération des données relatives à une personne peut accroître l'asymétrie du pouvoir de négociation entre les consommateurs et les entreprises qui leur vendent des produits et des services.

Dans le cadre d'un transfert de données d'une entité à une autre, l'organisation qui traite les données n'a souvent aucune relation directe avec l'entité qui les a recueillies dans un premier temps, et peut même se trouver à plusieurs niveaux de suppression. Il arrive que l'entité acquéreuse ne sache pas si la collecte et la communication des données ont été effectuées conformément aux lois sur la protection des données et de la vie privée.

Lorsque les données sont obtenues avec le consentement de la personne concernée (par exemple, les données relatives à l'utilisation des cartes de crédit, aux transactions financières, aux courriers électroniques), la question clé sera de savoir si le consentement a été obtenu de manière valide. S'agissant des données obtenues à partir d'espaces publics (par exemple, les données des satellites d'observation, les données recueillies par des drones, les images de surveillance, les données Dropcam), la question clé sera de savoir si les données ont réellement été obtenues à partir d'espaces publics, dans le respect des lois en matière de surveillance. Lorsque les données ont été obtenues sur Internet sans le consentement exprès de l'utilisateur (depuis un site Internet ou des API, documentées ou non), la question sera de savoir si les données ont été obtenues dans le cadre d'un accès autorisé. Des stratégies de certification pourraient voir le jour, afin de garantir que les données ont été soumises à des processus de désidentification, de pseudonymisation et d'anonymisation avant d'être commercialisées.

À l'heure actuelle, le marché des données est très fluide. Les entreprises achètent et vendent des données, et limitent leur risque de responsabilité, et donc le fardeau économique associé à la confidentialité des données, en obtenant des déclarations et des garanties contractuelles sur le respect des lois en matière de confidentialité, par exemple les lois relatives au consentement obligatoire de l'utilisateur. Des sociétés

telles que ZwillGen[159] conseillent les entreprises qui utilisent des mégadonnées sur la manière de gérer les risques économiques découlant de leur responsabilité dans le cadre des lois relatives au respect de la vie privée.

Tout cela ne rassure guère les personnes qui communiquent leurs données personnelles. Cette situation soulève également des questions sur la responsabilité des entités qui acquièrent des données en aval, notamment en ce qui concerne les niveaux de diligence voulue qu'elles doivent appliquer. La difficulté de suivre les opérations de traitement et de transfert des données exacerbe le problème de l'attribution de la responsabilité lorsque des données personnelles ont fait l'objet d'une utilisation non autorisée.

Les courtiers en données sont donc soumis à un examen de plus en plus minutieux, notamment en ce qui concerne les droits directs des usagers. Par exemple, dans son rapport sur la protection de la vie privée, la Commission fédérale du commerce des États-Unis désigne les courtiers en données comme des acteurs qui doivent permettre aux usagers d'accéder à leurs données par le biais d'une plate-forme commune facile à trouver et à utiliser. Elle a en outre contribué à l'élaboration d'une législation qui permettrait aux usagers d'accéder aux données détenues par les courtiers et leur donnerait le droit de les contester ou de les faire supprimer[160].

En mai 2018, le petit État américain du Vermont a été le premier à promulguer une loi relative aux courtiers en données et à la protection des usagers (en anglais: *An Act relating to data brokers and consumer protection*)[161]. Cette nouvelle loi réglemente les entreprises qui recueillent, vendent ou cèdent à des parties tierces, sous licence, les informations personnelles des résidents du Vermont avec lesquels elles n'ont pas de relation directe. Elle impose aux courtiers en données de s'enregistrer en tant que tels auprès des autorités, de divulguer toute information relative à leurs activités de collecte de données et de maintenir des mesures de protection des données. Tout manquement à cette obligation constitue une violation du droit du Vermont sur la protection des usagers, qui peut donner lieu à une action en justice de la part du Procureur général ou d'un particulier. Le nouveau *Consumer Privacy Act* (2018) de la Californie impose également des restrictions sur les transferts de données aux courtiers[162].

L'élaboration de lois s'appliquant aux courtiers en données promet de créer un nouveau domaine des droits des usagers, qui pourront accéder aux données qui sont détenues sur eux, rectifier celles qui sont incorrectes et obtenir réparation en cas de violation de leurs droits.

7 La phase de post-engagement: la redevabilité envers les usagers concernant les problèmes après les faits

Lorsque des systèmes décisionnels automatisés complexes fonctionnent sans intervention humaine, il est nécessaire de veiller à ce que les créateurs, les concepteurs, les fabricants, les opérateurs, les chargés de maintenance et les utilisateurs des algorithmes et des systèmes assument la responsabilité des éléments du processus qui relèvent de leurs fonctions respectives. Pour ce faire, il convient de garantir la transparence ou la traçabilité, c'est-à-dire de veiller à ce que les dispositifs décisionnels automatisés

---

[159] https://www.zwillgen.com/.
[160] Commission fédérale du commerce des États-Unis, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers*, mars 2012, p. 27. Disponible à l'adresse suivante: http://www.ftc.gov/os/2012/03/120326privacyreport.pdf.
[161] "An act relating to data brokers and consumer protection", projet de loi nº 764 de la Chambre du Vermont ("H-764"), disponible ici.
[162] Voir https://iapp.org/resources/article/california-consumer-privacy-act-of-2018/ et https://iapp.org/resources/topics/california-consumer-privacy-act/.

puissent justifier leurs décisions et notamment expliquer pourquoi ils privilégient une décision par rapport à d'autres possibilités. Cela nécessite de documenter chaque décision prise concernant les données sélectionnées, leur traitement et la conception des algorithmes. Enfin, les créateurs, concepteurs, fabricants, exploitants, chargés de maintenance et utilisateurs des algorithmes et des systèmes doivent assumer la responsabilité économique de leurs décisions, le cas échéant, sous la forme d'une obligation légale.

Cette section s'intéresse aux droits des usagers en cas de problème survenant après qu'ils ont partagé leurs données personnelles ou après que des données personnelles les concernant (communiquées par eux ou par autrui) ont été utilisées à leur détriment ou contre eux. La section 7.1 étudie dans un premier temps le droit des usagers d'intervenir face aux problèmes liés aux données qui pourraient être utilisées dans des décisions les concernant. Il s'agit notamment du droit de rectifier les données incorrectes détenues à leur sujet et de faire effacer certaines données. Cette question est abordée dans la section consacrée à l'étape de post-engagement, soit après qu'une entreprise a obtenu des données à caractère personnel, mais il est tout à fait possible qu'elle ne soit que le prélude à un autre engagement, au moment où ces données sont utilisées.

Cette section s'intéresse ensuite à la position de l'usager qui a été affecté par l'utilisation de ses données personnelles dans le cadre des mégadonnées et de l'apprentissage automatique (par exemple, suite à une décision ayant des conséquences juridiques ou similaires). La section 7.2 étudie les difficultés que les mégadonnées et l'apprentissage automatique posent dans le cadre des approches traditionnelles en matière de transparence et de responsabilité, notamment pour ce qui est du droit d'obtenir une explication au sujet des déductions et des décisions fondées sur ces technologies. La section 7.3 s'intéresse quant à elle aux droits de l'usager à contester les décisions qui ont été prises à son sujet à l'aide des processus liés aux mégadonnées et à l'apprentissage automatique. Enfin, la section 7.4 porte sur la question consistant à démontrer qu'il y a bien eu préjudice. Il ne peut pas y avoir de responsabilisation sans responsabilité.

7.1 Garantir les droits des usagers en matière d'accès, de rectification et d'effacement

L'une des principales garanties offertes aux usagers par les lois sur la protection des données et de la vie privée est le droit d'accéder aux données détenues par une organisation à leur sujet et de rectifier les erreurs qu'elles contiennent, ou de les compléter si elles sont incomplètes[163].

Par exemple, la loi californienne de 2018 sur la protection de la vie privée des consommateurs (*California Consumer Privacy Act*) oblige les entreprises qui recueillent des informations personnelles sur les résidents californiens à leur divulguer (gratuitement et sur demande) les types d'informations personnelles qu'elles ont recueillies sur eux au cours de l'année passée. Il peut s'agir de différents éléments d'information recueillis et des catégories de tierces parties avec lesquelles les informations ont été partagées[164]. Le RGPD européen confère aux individus le droit d'être informés

[163] D'après les lignes directrices de l'OCDE régissant la protection de la vie privée, "le droit des personnes physiques d'avoir accès aux données de caractère personnel et de les contester est, en règle générale, considéré comme étant peut-être la principale garantie de protection de la vie privée". *Lignes directrices de l'OCDE*, chapitre 3, 2013 (mémorandum explicatif des lignes directrices originales de 1980). En Europe, voir *Affaire C-131/12, Google Spain SL et Google Inc. contre Agencia Española de Protección de Datos (AEPD)*, EUR-Lex, 13 mai 2014. Voir Kelly, M. J. et Satola, D., "The Right to Be Forgotten". *University of Illinois Law Review*, vol. 1, 2017.
[164] "California Consumer Privacy Act" de 2018, CODE civil DE LA CALIFORNIE, paragraphes 178.110(a) et (b), et 178.130(a)(2).

de tout traitement éventuel de données personnelles les concernant, de recevoir une copie gratuite de ces données[165], de faire corriger les inexactitudes et de compléter les données incomplètes[166].

Ces droits sont également largement reconnus dans le droit international[167]. D'après les lignes directrices de l'OCDE régissant la protection de la vie privée, "le droit des personnes physiques d'avoir accès aux données de caractère personnel et de les contester est, en règle générale, considéré comme étant peut-être la principale garantie de protection de la vie privée".

Dans certaines juridictions, les personnes concernées peuvent avoir le droit d'accéder non seulement aux données communiquées et observées, mais aussi aux données déduites et dérivées (voir section 4.5). Il peut s'agir de profils élaborés par le responsable du contrôle des données, ainsi que d'informations sur la finalité du processus de traitement, sur les catégories de données détenues et sur leur origine[168].

Lorsque celles-ci sont vérifiables (date de naissance, adresse, niveau de salaire, état civil), les personnes concernées n'auront pas de mal à rectifier leurs données personnelles. Toutefois, dans le cas des mégadonnées et de l'apprentissage automatique, les données relatives à une personne peuvent comprendre des déductions plutôt que de simples faits sur sa vie.

Certaines déductions (niveaux de revenu, dépenses, maladies futures, âge du décès, etc.) peuvent être importantes pour prendre des décisions automatisées (ou humaines) concernant un individu (par exemple éligibilité aux services financiers ou prix de ces services). Certains suggèrent que le droit des personnes à rectifier leurs données personnelles ne devrait pas se limiter aux informations vérifiables parce que le caractère vérifiable d'une déduction ne détermine pas toujours son incidence sur l'individu concerné, et parce que ce dernier peut être en mesure de fournir des informations complémentaires (par exemple, des informations de santé à jour)[169].

Un nombre croissant de lois sur la protection des données confèrent aux individus le droit de faire effacer les données à caractère personnel les concernant lorsque celles-ci ne sont plus nécessaires aux fins pour lesquelles elles ont été recueillies ou traitées (également appelé droit à l'oubli)[170]. En vertu du RGPD, les

---

[165] RGPD de 2016, article 15.
[166] *Ibid*, article 16.
[167] L'observation générale 16 sur l'article 17 du Pacte international relatif aux droits civils et politiques prévoit que "chaque individu doit avoir le droit de réclamer l[a] rectification ou l[a] suppression" des fichiers contenant des données personnelles incorrectes. Comité des droits de l'homme, "Observation générale 16" (sur l'article 17 relatif au droit à la vie privée), supplément nº 40. Document officiel des Nations Unies A/43/40, 1988, paragraphe 10. Une observation générale sur une convention internationale est une remarque non contraignante visant à faciliter son interprétation. La Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (Convention 108) prévoit la "rectification ou l'effacement" de toute donnée traitée contrairement aux principes relatifs à la qualité des données, qui exigent que les données à caractère personnel faisant l'objet d'un traitement soient adéquates et à jour. Convention du Conseil de l'Europe pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (1981), article 8(c), avec référence à l'article 5. De même, en vertu du Cadre de protection de la vie privée de l'APEC, les individus devraient avoir le droit de "contester l'exactitude des informations les concernant et, si possible et selon le cas, de les faire rectifier, compléter, modifier ou supprimer". Cadre de protection de la vie privée de l'APEC de 2004, article 23(c).
[168] Groupe de travail "Article 29" sur la protection des données, "Guidelines on the Right to Data Portability", 16/EN WP 242 rev.01 10, 2017. Disponible à l'adresse suivante: https://ec.europa.eu/newsroom/article29/items/611233
[169] Groupe de travail "Article 29" sur la protection des données, "Avis 4/2007 sur le concept de données à caractère personnel", 01248/07/FR WP 136 (n 68); Groupe de travail "Article 29" sur la protection des données, "Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679", (n 19) 18.
[170] RGPD, article 17. Voir Kelly, M. J. et Satola, D., "The Right to Be Forgotten". *University of Illinois Law Review*, vol. 1, 2017. La nouvelle loi californienne sur la protection des consommateurs exige également que les entreprises suppriment les données personnelles

personnes ont le droit de faire effacer les données à caractère personnel les concernant lorsque ces informations ne sont plus nécessaires aux fins pour lesquelles elles ont été recueillies ou traitées et, si ledit traitement repose sur leur consentement, lorsqu'elles retirent leur consentement et qu'il n'existe aucun autre motif légal justifiant le traitement des données[171]. Pour reprendre un célèbre exemple, le droit à l'oubli a été exercé en Espagne contre Google[172]. La nouvelle loi californienne sur la protection des consommateurs exige également que les entreprises suppriment les données personnelles d'un consommateur lorsque celui-ci en fait la demande, à moins que ces informations ne lui soient nécessaires pour mener à bien certaines fonctions[173].

Reste encore à savoir si les déductions tirées de l'apprentissage automatique peuvent faire l'objet d'un droit d'accès, de rectification ou d'effacement. Cette question demeure incertaine dans de nombreux pays. Il est probable que les lois sur la protection des données de la plupart des pays seront appliquées de manière à donner plus de poids à l'intérêt d'une entreprise à conserver et à utiliser les données qu'elle a produites grâce à un processus de traitement reposant sur l'apprentissage automatique qu'aux intérêts des consommateurs en matière de vie privée, tout comme ses secrets commerciaux et sa propriété intellectuelle se verront attribuer une certaine valeur par rapport aux intérêts potentiellement nébuleux du consommateur[174]. Bien sûr, les données peuvent déjà avoir été partagées avec de tierces parties avant que le consommateur ne demande leur effacement, ce qui affaiblit encore ce type de recours.

À l'ère des mégadonnées, la prolifération des données personnelles entrave considérablement la capacité des personnes concernées à exercer ces droits.

---

d'un consommateur lorsque celui-ci en fait la demande, à moins que ces informations ne lui soient nécessaires pour mener à bien certaines fonctions. "California Consumer Privacy Act" de 2018, CODE civil de la Californie, paragraphes 178.105(a) et (b), et 178.130(a)(2).

[171] RGPD, article 17.

[172] *Affaire C-131/12, Google Spain SL et Google Inc. contre Agencia Española de Protección de Datos (AEPD)*, EUR-Lex, 13 mai 2014. Un ressortissant espagnol a déposé une plainte auprès l'Agence espagnole de protection des données (AEPD) au sujet d'articles sur Internet liant son nom à une procédure de saisie menée dans le cadre d'une vente aux enchères immobilière associée au recouvrement de dettes de sécurité sociale. M. Costeja González a demandé au journal de supprimer ou de modifier les pages en question, ou à Google Spain SL ou Google Inc. de supprimer ou de dissimuler les données personnelles figurant dans les résultats de recherche. Google s'est adressé à l'Audience nationale, le tribunal espagnol à compétence nationale, qui a sollicité la Cour de justice des Communautés européennes, laquelle a estimé que Google était l'un des responsables du contrôle des données à l'encontre duquel le droit à l'oubli pouvait être exercé et que, par conséquent, M. Costeja avait le droit de soumettre une telle demande et de la faire examiner par l'AEPD. Voir: Kelly, M. J. et Satola, D., "The Right to Be Forgotten". *University of Illinois Law Review*, vol. 1, 2017.

[173] *Ibid*, paragraphe 178.105.

[174] Voir, par exemple: Malgieri, G., "Trade Secrets v Personal Data: A Possible Solution for Balancing Rights". *International Data Privacy Law*, vol. 6, nº 102, 2016, p. 115.

7.2 Garantir la transparence et fournir des explications aux usagers

Principes FEAT de l'Autorité monétaire de Singapour

Responsabilité interne

7. L'utilisation des technologies d'AIDA pour prendre des décisions est approuvée par une autorité interne appropriée.

8. Les entreprises qui utilisent les technologies d'AIDA sont responsables des modèles d'AIDA élaborés en interne ou provenant de l'extérieur.

9. Les entreprises qui utilisent les technologies d'AIDA sensibilisent de manière proactive la direction et le conseil d'administration à leur utilisation.

Responsabilité externe

10. Les personnes disposent de canaux leur permettant de se renseigner sur les décisions les concernant prises grâce aux technologies d'AIDA, de faire appel et de demander leur révision.

Explication des décisions automatisées

L'obligation de rendre compte des décisions consiste généralement, dans un premier temps, à expliquer le fondement et la méthode décisionnelle, ou exige au moins une explication de ce type[175].

Certains préconisent de garantir aux usagers (comme l'ont fait plusieurs juridictions, telles que l'Union européenne) le droit à une explication lorsqu'une décision exclusivement automatisée (refus d'une demande de prêt, réduction de la limite de crédit, etc.) a des effets significatifs, notamment sur le plan juridique[176].

Cependant, fournir une explication aux usagers pose deux problèmes dans le contexte des mégadonnées et de l'apprentissage automatique:

Premièrement, les techniques sont difficiles à expliquer aux usagers, notamment dans un langage simple. Les modèles d'apprentissage automatique sont décrits comme "opaques"[177] ou comme des "boîtes noires"[178]. Même le fait de fournir le code source ne permettra pas aux informaticiens de savoir comment une décision a été prise, car l'apprentissage automatique est la science qui permet aux ordinateurs d'agir sans être explicitement programmés[179].

Deuxièmement, dans une certaine mesure, les modèles d'apprentissage automatique font l'objet de secrets commerciaux et de droits d'auteur sur les logiciels qui sont le résultat d'investissements et existent sur un marché commercial concurrentiel. L'opérateur d'un modèle d'apprentissage automatique peut être réticent à partager le codage ou l'explication de l'algorithme, de peur d'affaiblir la concurrence et de compromettre l'investissement initial.

Pour ce qui est de l'utilisation des algorithmes, ces facteurs posent des difficultés importantes en matière de responsabilité envers les usagers[180]. En particulier, la difficulté d'expliquer à un usager la relation entre les données d'entrée et les données de sortie empêche celui-ci de contester les décisions prises à son sujet. Néanmoins, même si des explications sont actuellement difficiles à produire, il se pourrait

[175] Doshi-Velez, F., *et al.*, "Accountability of AI Under the Law: The Role of Explanation", arXiv prétiré arXiv:1711.01134, 2017.
[176] Voir *Ethically Aligned Design*, à la note de bas de page 223, p. 160.
[177] Burrell, J., "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms". *Big Data & Society*, 2016.
[178] Pasquale, F., *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press, 2015.
[179] Stanford University, "Machine Learning", Coursera. Disponible à l'adresse suivante: https://www.coursera.org/learn/machine-learning/home/info [https://perma.cc/L7KF-CDY4].
[180] Voir "Accountable Algorithms", à la note de bas de page 128. Ohm, P. et Lehr, D., "Playing with the Data: What Legal Scholars Should Learn About Machine Learning". *University of California Davis Law Review*, 2017. Disponible à l'adresse suivante: https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

bien que l'on mette en œuvre les efforts nécessaires seulement une fois que les droits juridiques susvisés auront vu le jour.

D'importantes raisons peuvent justifier de tels efforts. L'acceptation, par l'ensemble de la société, des mégadonnées et de l'apprentissage automatique, en particulier la prise de décisions automatisée et les services qui en dépendent, dépendra au moins en partie de la confiance – en une utilisation raisonnable d'informations pertinentes. Il est communément admis qu'en matière d'apprentissage automatique, la corrélation et la prédiction sont des principes directeurs, et que la causalité et le raisonnement sont sans importance. En 2008, Chris Anderson a déclaré que la méthode scientifique était obsolète, dépassée par le pouvoir corroborant des corrélations de masse [181] . L'apprentissage automatique identifie les corrélations existant entre plusieurs facteurs (à ne pas confondre avec les liens de causalité). Il est capable de faire des prédictions sur le comportement futur, mais pas de justifier ces prédictions.

L'apprentissage automatique consiste à exposer un système informatique à de grands volumes de données (provenant d'exemples passés), à l'entraîner à observer les formes qui s'en dégagent et à déduire une règle à partir de ces formes. Plutôt que d'établir directement des règles, les êtres humains génèrent un processus informatisé de définition de règles. Cette abstraction, ou déconnexion, entre les êtres humains et le processus décisionnel, complique la vérification des règles établies. Il est donc difficile de leur demander des comptes lorsque les règles ou leurs résultats ne répondent pas aux objectifs politiques, ou vont même à l'encontre des lois en vigueur, notamment en matière de discrimination. En effet, non seulement les personnes ordinaires ne comprennent pas les modèles d'apprentissage automatique, mais souvent, même ceux qui les ont élaborés sont incapables d'expliquer pourquoi ils sont efficaces.

Toutefois, dans de nombreux secteurs, il est inacceptable dans la pratique que les modèles d'apprentissage automatique soient compris seulement par les spécialistes des données et les programmeurs informatiques. Dans le secteur médical, bancaire ou des assurances, par exemple, les chercheurs et même les praticiens doivent comprendre les modèles d'apprentissage automatique sur lesquels ils s'appuient s'ils veulent avoir confiance en eux et en leurs résultats. Il peut s'avérer nécessaire de trouver un compromis entre, d'une part, le maintien de la transparence et de l'interprétabilité des modèles et des processus de modélisation (ce qui suppose de réduire la complexité autant que possible) et, d'autre part, le développement de modèles d'apprentissage automatique évolutifs à des fins de précision et d'efficacité (ce qui les rend plus complexes et plus difficiles à expliquer).

En outre, la précision de l'apprentissage automatique dépend de la manière dont les données utilisées pour entraîner et valider les modèles d'apprentissage automatique sont sélectionnées et conservées. Elle dépend également de la bonne articulation de la tâche du modèle, de la possibilité de formuler des hypothèses bien définies et de la sélection de paramètres d'efficacité pertinents. En fin de compte, si l'on dispose de suffisamment de temps et de ressources, un programme informatique doit pouvoir être expliqué, sinon il n'y a aucune raison d'avoir confiance dans l'exactitude de ses conclusions[182].

Alors que certains avancent que la complexité défie toute explication, d'autres affirment qu'une telle vision dissimule la facilité de compréhension des algorithmes, et que plutôt que d'écarter les systèmes

---

[181] Anderson, C., "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete". *Wired*, 23 juin 2008. Disponible à l'adresse suivante: https://www.wired.com/2008/06/pb-theory/.

[182] Hildebrandt, M., dir. Bayamlıoğlu, E., Baraliuc, I., Janssens, L. et Hildebrandt, M., "Preregistration of machine learning research design. Against P-hacking", dans *Being Profiled: Cogitas Ergo Sum*. Amsterdam University Press, 27 septembre 2018 (à paraître). Disponible sur SSRN, à l'adresse suivante: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3256146.

qui produisent de mauvais résultats en les considérant comme fondamentalement impénétrables et donc incontrôlables, nous devrions simplement qualifier l'application d'une technologie inadéquate pour ce qu'elle est: une faute professionnelle, commise par le responsable du contrôle d'un système[183]. Il n'en reste pas moins qu'il est difficile de fournir des explications sur les décisions automatisées qui puissent être facilement comprises par des individus inexpérimentés.

Une réglementation qui permet des explications adéquates

Lorsqu'un fournisseur de services financiers prend une décision en fonction de données d'entrée (par exemple, les niveaux de revenu, le patrimoine, le code postal), celle-ci repose en fin de compte sur les déductions faites à partir de ces sources d'information, par exemple si le risque de défaut de paiement de l'individu pour un prêt d'un certain montant sur une certaine période est trop élevé pour justifier le prêt. En général, les lois sur la protection des données ne prévoient pas de dispositions contre les déductions déraisonnables, laissant ces questions à des lois sectorielles spécifiques, à supposer qu'elles existent. En effet, la plupart des lois sur la protection des données n'exigent pas du responsable du contrôle des données qu'il fournisse une explication pour chaque décision automatisée qui a été prise. Tout au plus, elles exigent généralement d'avertir les personnes concernées qu'une décision future sera automatisée, et éventuellement, de leur donner la possibilité de s'y opposer[184].

Certains pays vont un peu plus loin. Par exemple, la loi brésilienne de 2018 sur la protection des données donne au consommateur le droit de demander un réexamen des décisions prises uniquement sur la base d'un traitement automatisé de données à caractère personnel qui affectent ses intérêts. Il peut notamment s'agir des décisions visant à définir son profil ou à évaluer certains aspects de sa personnalité, et du droit de demander des informations claires et pertinentes au sujet des critères et des procédures sur lesquels repose une décision automatisée[185].

Certains décideurs politiques penchent pour un examen plus approfondi des décisions automatisées dans le cadre de la législation sur la protection des données et de la vie privée. Le Groupe de travail "Article 29"de l'Union européenne sur la protection des données, par exemple, a conseillé aux responsables du contrôle des données d'éviter de trop se fier aux corrélations et de fournir des informations significatives à la personne concernée sur la logique impliquée dans la prise de décisions automatisée[186] (à savoir, par exemple, les principales caractéristiques prises en compte pour prendre telle ou telle décision, l'origine de ces informations, leur pertinence). Dans le même ordre d'idées, les responsables du contrôle des données peuvent être tenus de démontrer que leurs modèles sont fiables en vérifiant leur précision statistique et en corrigeant les inexactitudes, notamment pour éviter de prendre des décisions discriminatoires[187].

D'après le Future of Privacy Forum, l'explication des modèles d'apprentissage automatique doit renseigner sur la manière dont le modèle a été choisi, analyse juridique et technique à l'appui. Il s'agirait

---

[183] Kroll, J. A., "The fallacy of inscrutability". *Philosophical Transactions of the Royal Society A*, vol. 376, 20180084, 2018. Disponible à l'adresse suivante: https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0084.
[184] Wachter, S., Mittelstadt, B. et Floridi, L., "Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation". *International Data Privacy Law*, 2017. Disponible à l'adresse suivante: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469.
[185] Article 22.
[186] RGPD, articles 13 à 15.
[187] Groupe de travail "Article 29" sur la protection des données, "Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679", voir la note de bas de page 55, p. 28-29.

notamment de déterminer les compromis à effectuer entre explicabilité et précision. Cette démarche permettrait d'enregistrer les décisions visant à complexifier un modèle au détriment de l'explicabilité, et de tenir compte de la matérialité des résultats pour les individus et les tierces parties (par exemple, l'enjeu d'un traitement médical est plus important que celui de recommandations cinématographiques)[188].

Certains soutiennent que l'absence d'explications efficaces compromet le principe de responsabilité et que les lois sur la protection des données et de la vie privée doivent conférer aux usagers un droit effectif à des déductions raisonnables[189].

Lorsque les déductions risquent fort de rendre des décisions défavorables, de nuire à la réputation ou de porter atteinte à la vie privée, un tel droit pourrait exiger du responsable du contrôle des données qu'il explique avant le traitement (*ex ante*) la pertinence de certaines données pour les déductions à tirer, la pertinence des déductions pour le type de décision à prendre et le traitement automatisé, ainsi que la précision et la fiabilité statistique de la méthode utilisée. Ces explications pourraient s'accompagner de la possibilité de contester les décisions une fois qu'elles ont été prises (*ex post*).

Cela permettrait, en plus de contester une décision automatisée sur la base de l'exactitude des données utilisées, de contester les déductions vérifiables sur lesquelles elle repose (niveau de revenu, patrimoine, état de santé, état civil, etc.). Les déductions non vérifiables peuvent être remises en question par des données supplémentaires, qui pourraient venir modifier les conclusions.

Les efforts visant à introduire une réglementation qui s'immisce dans le fond des décisions ou le processus décisionnel, par opposition aux simples processus de collecte, d'utilisation et de partage des données, peuvent être considérés par certains comme un fardeau pour un secteur novateur émergent qui devrait être laissé libre de développer des produits qui profitent aux usagers et de les affiner sous la pression de la concurrence. D'autres y voient une tentative de rééquilibrer la marginalisation des usagers résultant de la suppression des éléments humains des principales étapes de la prise de décisions (voir section 7.3). Une interaction humaine donne la possibilité de rencontrer un décideur ou quelqu'un qui peut influencer le décideur, d'échanger avec lui, et de lui montrer où les déductions étaient erronées. Pour que le droit à l'intervention humaine dans les décisions automatisées ait de la substance, il peut être nécessaire de préciser le degré d'intégrité que ladite intervention cherche à atteindre dans le cadre du processus.

Les lois sur la protection des données ne garantissent généralement pas la précision de la prise de décisions, et il est probable qu'il en soit de même pour l'exactitude des données d'inférence, de sorte que même lorsque des déductions incorrectes ont été tirées de données exactes, l'individu pourrait ne pas avoir le droit de les rectifier[190].

Cela relèverait plus généralement de lois sectorielles, telles que la loi sur les services financiers, mais dans la plupart des pays, ces lois interdisent uniquement la prise de décisions discriminatoire en fonction de critères spécifiques (tels que l'appartenance ethnique, le genre ou la religion) et ne prescrivent pas la justesse de la décision elle-même. En ce sens, un mauvais algorithme s'apparente à un mauvais employé

[188] Future of Privacy Forum, *Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models*, 2018.
[189] Voir Wachter, S. et Mittelstadt, B., à la note de bas de page 56.
[190] Voir Wachter, S. et Mittelstadt, B., à la note de bas de page 56.

de banque qui ne parvient pas à prendre une bonne décision en raison d'un manque de jugement ou de son inexpérience: il peut s'agir d'une mauvaise pratique commerciale, mais ce n'est pas illégal.

Toutefois, une loi sur les services financiers peut prescrire certaines procédures en vue de favoriser les bonnes décisions. Par exemple, elle peut exiger d'un fournisseur de services financiers qu'il procède à une évaluation des besoins de son client afin de lui proposer un produit qui soit plus susceptible de lui convenir[191]. Elle pourrait également exiger des évaluations des risques qui garantiront la prise en compte des risques, y compris dans les algorithmes eux-mêmes.

Clarification des explications

Il a également été suggéré de remplacer ou de compléter les explications par un retour d'information "contrefactuel", aussi bien positif que négatif, sur les décisions automatisées (et seulement sur celles qui sont en grande partie automatisées). Ces explications contrefactuelles peuvent informer la personne concernée non pas tant sur la manière dont une décision a été prise que sur les variations des données d'entrée qui auraient pu conduire à une décision différente[192]. Par exemple, un fournisseur de services financiers numériques pourrait informer le consommateur de la manière qui suit: "Votre demande de prêt indique que votre revenu annuel est de 30 000 dollars. S'il avait été de 45 000 dollars, on vous aurait proposé un prêt[193]."

Bien sûr, il existe de nombreuses variables d'entrée dans la prise de décisions, et de nombreuses combinaisons qui pourraient produire un nombre quasi infini d'éléments contrefactuels possibles. Il est donc peu probable que l'on puisse réduire l'explication d'une décision à une, voire même à plusieurs, variables. En outre, avec une telle approche, les institutions doivent faire attention lorsqu'elles s'engagent à fournir un service dans des conditions alternatives (pour reprendre l'exemple précédent, si l'individu revient avec un revenu de 45 000 dollars, il pourrait être en droit de s'attendre à ce que sa demande de prêt soit approuvée).

Toutefois, si ces éléments contrefactuels étaient codés dans le service, les résultats contrefactuels pourraient être rapidement communiqués au consommateur, qui pourrait éventuellement expérimenter différents niveaux de variables. En effet, les interfaces consommateurs pourraient même proposer une échelle mobile de données d'entrée, ce qui permettrait au consommateur une certaine expérimentation. Il peut donc être possible de fournir des éléments contrefactuels qui amélioreraient la compréhension du consommateur et lui offriraient la possibilité de contester une décision, voire de modifier sa situation pour obtenir une décision plus favorable. Par exemple, lorsqu'il comprend qu'arrêter de fumer lui donne droit à une assurance maladie, ou que rembourser telle ou telle dette ou augmenter ses revenus entraîne

---

[191] Par exemple, les directives sur la protection des consommateurs publiées par la Banque centrale du Kenya, (*Guideline on Consumer Protection*, section 3.2.1(c)(iv)) exigent que les banques ne profitent pas d'un consommateur qui n'est pas en mesure de comprendre le caractère ou la nature d'une transaction proposée. Une banque doit donc s'enquérir des besoins spécifiques du consommateur afin de lui proposer des produits ou des services adaptés à ceux-ci. La section 3.2.2(i) de ces lignes directrices dispose quant à elle que, selon la nature de la transaction et sur la base des informations communiquées par un client, une banque doit évaluer et comprendre les besoins du client avant de lui proposer un service. La section 3.2.4(a)(ii) exige par ailleurs que les banques, lorsqu'elles prodiguent des conseils à leurs clients, s'assurent que tout produit ou service recommandé à un client est adapté à ce dernier.

[192] Pasquale, F., *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2015; Mayer-Schönberger, V. et Ramge, T., *Reinventing Capitalism in the Age of Big Data*, Basic Books, 2018; Wachter, S., Mittelstadt, B. et Russell, C., "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR", arXiv prétiré, 2017, disponible à l'adresse suivante: https://arxiv.org/abs/1711.00399; *Accountable Algorithms*, à la note de bas de page 128.

[193] Wachter, S., Mittelstadt, B. et Russell, C., "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR". *Harvard Journal of Law & Technology*, 2018. Disponible à l'adresse suivante: https://arxiv.org/abs/1711.00399.

une décision de crédit positive, l'individu concerné peut exercer un plus grand contrôle sur sa vie qu'en étant le sujet passif d'une décision.

Cette approche peut réduire l'écart entre les positions de négociation et aboutir à la proposition et à l'acceptation d'une offre commercialement rentable pour un service souhaité, une situation qui bénéficie à la fois au fournisseur et au consommateur. Il y a donc de bonnes raisons d'attendre des acteurs du marché qu'ils introduisent de tels paramètres pour différencier leurs services sur un marché concurrentiel, bien qu'un "coup de pouce" réglementaire puisse parfois favoriser l'adoption et la généralisation de telles pratiques.

L'approche contrefactuelle pourrait en outre atténuer les craintes selon lesquelles le fait d'exiger des explications amène à révéler des secrets commerciaux et à enfreindre les obligations de non-divulgation. En fournissant des éléments contrefactuels, on peut au contraire éviter d'avoir à divulguer la logique interne des algorithmes du système décisionnel. Il s'agit d'une stratégie de transparence concrète et axée sur les résultats, qui peut présenter des avantages par rapport à l'obligation de fournir des explications, qui sont parfois si complexes qu'elles n'améliorent ni la compréhension ni la situation du consommateur.

Si faire référence à des éléments contrefactuels est une stratégie relativement subtile pour améliorer la position des consommateurs, notamment en leur donnant de nouveaux moyens d'obtenir les services recherchés, il existe des méthodes plus approfondies pour améliorer la responsabilité des systèmes d'apprentissage automatique. Il pourrait être possible, par exemple, d'examiner et de certifier les propriétés des systèmes informatiques, et de veiller à ce que les décisions automatisées soient prises conformément aux règles convenues, notamment à des fins de protection contre la discrimination. Certains qualifient cette approche de "régularité procédurale"[194].

Pour qu'un modèle d'apprentissage automatique fonctionne de manière responsable, le principe de responsabilité doit être incorporé à l'étape de conception du système. Les concepteurs de systèmes, et ceux qui supervisent la conception, doivent considérer en priorité les principes de responsabilité et de surveillance. L'Initiative mondiale de l'IEEE pour les considérations éthiques dans l'intelligence artificielle et les systèmes autonomes recommande ce qui suit[195]:

> *Bien que l'on reconnaisse que cela ne peut être fait à l'heure actuelle, les systèmes d'intelligence artificielle devraient être conçus de manière à pouvoir, lorsqu'on leur demande, montrer à l'utilisateur humain le processus enregistré qui a conduit à telle ou telle action, repérer dans la mesure du possible les sources d'incertitude et énoncer toutes les hypothèses sur lesquelles ils s'appuient.*

L'IEEE propose également de concevoir et de programmer les systèmes d'intelligence artificielle dans une optique fondamentale de transparence et de responsabilité[196], ainsi que d'informer de manière proactive les utilisateurs de leur incertitude[197].

---

[194] Voir "Accountable Algorithms", à la note de bas de page 128. Ohm, P. et Lehr, D., "Playing with the Data: What Legal Scholars Should Learn About Machine Learning". *University of California Davis Law Review*, 2017. Disponible à l'adresse suivante: https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.
[195] Voir *Ethically Aligned Design*, à la note de bas de page 223, p. 159.
[196] *Ibid*, p. 152.
[197] *Ibid*, p. 159.

7.3 Donner aux usagers les moyens de contester les décisions prises

Comme nous l'avons vu à la section 7.2, les lois sur la protection des données ne donnent généralement pas le droit de contester l'exactitude des décisions prises en fonction des données communiquées. Toutefois, les usagers ont de plus en plus souvent la possibilité de contester des décisions prises exclusivement dans le cadre d'un traitement automatisé. De nouveaux risques découlent de la prise de décisions automatisée dans les domaines des services financiers qui affectent la vie des personnes, tels que le crédit, les assurances et les produits financiers risqués ou coûteux[198]. L'Initiative mondiale de l'IEEE recommande que les personnes disposent d'un forum pour faire valoir les circonstances atténuantes que le système d'intelligence artificielle pourrait ne pas prendre en compte – à savoir un recours à une intervention humaine[199]. Un nombre croissant de lois sur la protection des données et de la vie privée, dont le RGPD, prévoient le droit d'obtenir une intervention humaine, d'exprimer son point de vue et de contester les décisions prises[200].

Ce droit relève des notions de procédure régulière, qui peuvent être compromises si les décisions sont prises par un ordinateur sans autre recours. Il part également de l'idée que traiter les personnes avec respect et dignité implique de veiller à ce que les décisions importantes qui affectent leur vie ne soient pas prises par un simple ordinateur, mais par un autre être humain. Cette préoccupation est exacerbée par le risque que les ordinateurs produisent des résultats erronés ou discriminatoires[201].

La possibilité de contester une décision automatisée ne consiste pas simplement à cliquer sur une demande de réexamen et à recevoir une autre décision automatisée définitive, ce qui ne ferait que produire une autre décision automatisée soumise à un droit de contestation. En fin de compte, si une décision automatisée doit être révisée, il est nécessaire de s'assurer qu'elle est soumise à une certaine forme d'intervention humaine, dans le cadre de laquelle la personne concernée a la possibilité de présenter son point de vue à un autre être humain, qui déterminera si la décision automatisée doit être révisée.

Ladite intervention humaine peut varier dans son degré d'implication, allant d'un plein droit d'appel sur l'ensemble de l'affaire, à une simple vérification de l'algorithme, en vue de s'assurer que les données d'entrée sont exactes, mais pas d'évaluer son fonctionnement. Dans l'ensemble, il est toutefois probable que les droits de contestation des décisions prises dans le cadre d'une intervention humaine se limiteront aux cas de violation des principes de protection des données (les données d'entrée sont incorrectes ou incomplètes, le consentement requis de la part de la personne concernée n'a pas été obtenu, etc.). Il s'agirait donc davantage de questions de procédure que de questions de fond. Le "raisonnement" qui sous-tend le fond des décisions et qui oriente la conception et le fonctionnement des algorithmes ne serait probablement pas sujet à contestation en vertu des lois sur la protection des données.

Cette situation ne signifie pas que les lois, réglementations et normes sectorielles ne peuvent pas exiger des fournisseurs qu'ils modifient ou annulent leurs décisions lorsqu'elles sont générées par des modèles d'apprentissage automatique pour des raisons de fond, mais plutôt qu'en l'absence de telles lois,

---

[198] Groupe de travail "Article 29" sur la protection des données, "Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679", voir la note de bas de page 55, p. 10.
[199] Initiative mondiale de l'IEEE (voir note de bas de page 223), p. 153.
[200] RGPD, article 22.
[201] Roth, A., "Trial by Machine". *Georgetown Law Journal*, vol. 104, 2016, p. 1245.

réglementations ou normes, les consommateurs disposent d'un recours limité pour contester une décision automatisée[202].

Alors que les personnes peuvent être protégées contre la collecte, l'utilisation et le partage prescrits de leurs données personnelles (en particulier les données sensibles ou appartenant à des catégories particulières), ainsi que contre l'utilisation de données inexactes ou incomplètes dans la prise de décisions automatisée les concernant, elles sont peu protégées lorsqu'il s'agit de la manière dont les décisions sont effectivement prises.

7.4 Évaluer les préjudices et la responsabilité à l'égard des usagers

Le principe de responsabilité repose en définitive sur le fait d'être tenu pour responsable devant la loi, y compris en matière de réparation des préjudices causés. L'une des difficultés de l'élaboration de politiques, d'obligations légales et de recours en faveur des usagers dans le domaine de la protection des données tient à la nature intangible du préjudice contre lequel les usagers doivent être protégés ou pour lequel ils doivent être indemnisés.

Cette difficulté peut compromettre la réclamation d'un usager dès le départ. Pour avoir la possibilité de soumettre une demande d'indemnisation au tribunal, il est généralement nécessaire d'affirmer que l'on a subi un préjudice. Les tribunaux ont rencontré des difficultés à identifier les préjudices causés en cas de non-respect des lois sur la protection des données et de la vie privée, parvenant souvent à des points de vue juridiques très différents. De nombreuses réclamations ont été rejetées parce que les usagers n'ont pas réussi à démontrer le préjudice qu'ils avaient subi.

La question de savoir si une personne a subi ou non un préjudice est souvent examinée au regard d'un élément contrefactuel. Il s'agira donc de déterminer si la personne se trouve dans une situation plus inconfortable que si l'événement ne s'était pas produit[203]. Il est particulièrement difficile de démontrer un préjudice lorsqu'il n'y a pas encore eu de perte pécuniaire ou physique, par exemple lorsqu'il y a eu violation d'un système et que des données ont été obtenues sans autorisation, mais qu'elles n'ont pas (encore) été utilisées pour voler de l'argent. Le préjudice peut être considéré comme conjectural, alors que dans certains systèmes juridiques, les plaignants doivent démontrer qu'ils ont effectivement subi un préjudice[204].

Les théories du préjudice causé par l'obtention illégale de données à caractère personnel incluent notamment le risque de fraude ou d'usurpation d'identité, et l'anxiété que l'individu peut éprouver face à ces risques. Si les préjudices immatériels sont plus difficiles à reconnaître et à évaluer, ils peuvent être tout aussi réels et concrets que les dommages pécuniaires[205]. En effet, non seulement les préjudices immatériels peuvent être bien réels, mais il est de plus en plus souvent admis que le risque même de préjudice – c'est-à-dire lorsque le dommage ne s'est pas encore matérialisé, mais que le risque existe – devrait être considéré comme un préjudice légitime aux fins des réclamations des usagers.

---

[202] Voir Wachter, S. et Mittelstadt, B., à la note de bas de page 56.
[203] Feinberg, J., "Wrongful Life and the Counterfactual Element in Harming", dans *Freedom and Fulfillment*, 1992, p. 3.
[204] Par exemple, dans l'affaire opposant Clapper à Amnesty International USA, 133 S. Ct. 1138 (2013), la Cour suprême des États-Unis a rejeté les réclamations formulées à l'encontre du gouvernement américain en raison de la collecte accrue de données à des fins de surveillance, au motif que les plaignants n'avaient pas démontré de "préjudice de fait".
[205] Spokeo, *ibid.*

De tels préjudices peuvent être évalués en fonction de la probabilité et de l'ampleur du préjudice éventuel, de la sensibilité des données exposées, de la possibilité d'atténuer le préjudice et du caractère raisonnable des mesures préventives[206]. Les tribunaux avaient tendance à être plus favorables aux plaignants en cas d'usurpation d'identité, en raison du risque de fraude[207], ou de publication d'informations inexactes à leur sujet[208].

Il existe plusieurs types de préjudices potentiels dans le cadre d'une prise de décisions automatisée[209]. Les décisions de ce type peuvent faire subir une perte économique à une personne. On peut par exemple refuser un bien ou un service à un individu (ou les lui proposer à un prix plus élevé) parce qu'il a été classé dans groupe particulier (notamment en raison du quartier dans lequel il réside, une pratique parfois qualifiée de "*redlining*"). Les personnes qui souhaitent obtenir un prêt, demandent une augmentation de la limite de crédit ou envisagent de souscrire un contrat d'assurance peuvent subir une perte d'opportunité lorsqu'elles sont filtrées en fonction de leur appartenance ethnique, ou de leurs informations génétiques ou de santé.

Dans les pays où ils impliquent une discrimination liée à l'appartenance ethnique, à la religion, au casier judiciaire ou à la santé, certains préjudices sont illégaux. Dans ces cas, les lois en vigueur protégeront spécifiquement certaines catégories de personnes et pourront interdire les résultats discriminatoires. Cependant, lorsque l'appartenance à une catégorie protégée n'entre pas en jeu, il peut s'avérer difficile de démontrer un préjudice.

Une autre difficulté rencontrée par les usagers lésés par les systèmes reposant sur les mégadonnées et l'apprentissage automatique est d'identifier qui doit être tenu responsable des préjudices causés: l'entreprise qui emploie le système, celle qui a codé les algorithmes ou celle qui a fourni les données? Il peut être impossible, pour l'usager, de démontrer la cause précise d'un préjudice et d'identifier la partie responsable.

La section 6.2 a abordé les différentes mesures que les opérateurs de systèmes d'apprentissage automatique pouvaient prendre pour réduire le risque de biais. En outre, certains ont suggéré d'exiger de certaines entreprises ayant recours à l'intelligence artificielle et à l'apprentissage automatique qu'elles souscrivent une assurance ou d'autres garanties de responsabilité financière, afin de fournir un moyen de recours aux personnes lésées[210]. Bien que cela puisse paraître plus évident en cas de dommages corporels impliquant des équipements tels que des véhicules autonomes qu'en cas de réclamation pour perte d'opportunité, une telle approche pourrait être envisagée lorsque les préjudices découlent d'une violation des données par des processeurs traitant de grands ensembles de données.

---

[206] Solove, D. J. et Citron, D., "Risk and Anxiety: A Theory of Data Breach Harms". *Texas Law Review*, vol. 96, 2018, p. 737.
[207] Dans l'affaire Remijas c. Neiman Marcus Group, LLC (794 F.3d 688, 693-694, Septième circuit, 2015), la Cour fédérale des États-Unis a estimé que le fait que les plaignants savaient que les informations de leur carte de crédit personnelle avaient été volées par des personnes qui prévoyaient de les utiliser à mauvais escient (les cartes d'autres plaignants ayant fait l'objet d'une utilisation frauduleuse) constituait un préjudice suffisant pour leur donner le droit d'entamer des procédures judiciaires.
[208] Dans l'affaire Spokeo (voir note de bas de page 114), la Cour suprême des États-Unis a estimé que lorsqu'un moteur de recherche de personnes décrivait un individu de manière incorrecte, cela pouvait entraîner un risque de préjudice suffisamment grave pour lui permettre d'intenter un procès.
[209] Pour une vibrante description de ces types de préjudices, voir O'Neill, C., *Weapons of Math Destruction*, 2016. Pour une taxonomie utile des préjudices que pourraient causer les processus décisionnels automatisés, voir Future of Privacy Forum, *Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making*, décembre 2017.
[210] Initiative mondiale de l'IEEE (voir note de bas de page 223), p. 156.

Il a également été suggéré aux tribunaux et aux législateurs de s'inspirer du riche corpus de lois sur la responsabilité du fait des produits lorsqu'ils traitent des réclamations pour une certaine forme de préjudice résultant de l'intelligence artificielle et de l'apprentissage automatique. Dans certains cas, cela pourrait revenir à appliquer une responsabilité stricte, c'est-à-dire sans qu'il soit nécessaire de démontrer un lien de causalité, un acte de négligence ou une faute (et encore moins une intention). Là encore, les dispositifs de recours devraient inciter les fournisseurs à s'attaquer aux problèmes avant et après leur apparition. Par exemple, la législation sur la responsabilité du fait des produits cherche souvent à éviter de compromettre les mesures visant à inciter fabricants à corriger les défauts de leurs produits à la suite d'un préjudice, de peur que cela ne soit considéré comme une reconnaissance de la responsabilité du préjudice. Dans ce cas, la loi dispose que ces démarches ne sont pas admissibles comme preuve de la faute[211].

Dans l'ensemble, il reste beaucoup à faire dans la plupart des juridictions pour offrir aux usagers des recours efficaces en cas d'atteinte à leur vie privée et de risques liés aux mégadonnées et à l'apprentissage automatique.

8 Gestion des risques, conception et éthique

Les sections précédentes ont abordé la protection des usagers et la confidentialité des données, en mettant l'accent sur le processus de traitement et les recours juridiques et réglementaires. L'incertitude qui entoure ces questions représente un risque pour les entreprises, qui pourraient être tenues responsables du non-respect des lois contre la discrimination ou encourir une lourde responsabilité en matière de dommages et intérêts en cas d'atteinte à la vie privée et à la sécurité des données. La présente section examine les différentes mesures que les entreprises peuvent prendre pour atténuer ces risques.

8.1 Assurer la gestion des risques

Une approche courante en cas d'incertitude est d'appliquer des cadres et des processus de gestion des risques. Une bonne conception de modèle de mégadonnées incorpore donc la gestion des risques[212]. Par exemple, certains fournisseurs de services financiers, comme Mastercard, appliqueront un processus intersectoriel d'exploration de données (*Cross-Industry Standard Process for Data Mining*, ou CRISP/DM), qui fournit une approche structurée de la planification des projets d'exploration de données[213].

Ces cadres et processus peuvent être utilisés pour évaluer les risques liés à la vie privée et à la discrimination des usagers, comme tout autre risque. Le National Institute of Standards and Technology des États-Unis a récemment entrepris d'élaborer un cadre pour la protection de la vie privée[214], axé sur des approches de gestion des risques calquées sur son cadre de cybersécurité. Ce cadre souligne l'importance d'accorder la priorité à la gestion des risques plutôt qu'à des stratégies de conformité visant à répondre au strict minimum.

Les processus de gestion des risques relatifs aux systèmes d'apprentissage automatique peuvent inclure la documentation des objectifs et des hypothèses formulées, et l'utilisation des "trois lignes de défense"

---

[211] Initiative mondiale de l'IEEE (voir note de bas de page 223), p. 156.
[212] *Ibid.*
[213] Voir https://www.sv-europe.com/crisp-dm-methodology/.
[214] Voir https://www.nist.gov/privacy-framework.

qui garantissent la séparation des éléments suivants (par processus, rôle, partie prenante et mesure incitative):

- élaboration et expérimentation d'un modèle d'apprentissage automatique;
- validation et examen juridique dudit modèle; et
- vérification périodique tout au long de son cycle de vie[215].

Le suivi, l'amélioration et la responsabilisation continus des systèmes d'apprentissage automatique sont tributaires de la documentation de ces objectifs[216].

La gestion des risques peut s'appliquer aux données d'entrée et de sortie des modèles d'apprentissage automatique[217]:

En ce qui concerne les ***données d'entrée***, l'atténuation des risques commencera par la consignation des prérequis du modèle (par exemple, le degré d'actualité des données, leurs caractéristiques et leurs utilisations), le degré de dépendance des systèmes environnants à l'égard des données, la manière dont les données personnelles sont prises en compte – et à quelles fins – et dont elles sont protégées (par chiffrement ou autre), ainsi que leur traçabilité. La documentation permet une révision et une maintenance efficaces. Il s'agira notamment d'évaluer l'exhaustivité, l'exactitude, la cohérence,

[215] Voir, par exemple: Conseil des gouverneurs du Système fédéral de réserve et Bureau du Contrôleur de la monnaie, "Supervisory Guidance on Model Risk Management", avril 2011, disponible à l'adresse suivante: https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf; et les cadres européens, la Directive 2013/36/UE du Parlement européen et du Conseil du 26 juin 2013 concernant l'accès à l'activité des établissements de crédit et la surveillance prudentielle des établissements de crédit et des entreprises d'investissement; le Règlement nº 575/2013 du Parlement européen et du Conseil du 26 juin 2013 concernant les exigences prudentielles applicables aux établissements de crédit et aux entreprises d'investissement; et le guide de la Banque centrale européenne sur l'examen ciblé des modèles internes (*Targeted Review of Internal Models*, ou TRIM).
[216] Ainsi, l'Initiative mondiale de l'IEEE (voir note de bas de page 223) recommande que les systèmes automatisés génèrent des pistes de vérification enregistrant les faits et la législation en appui à la prise de décisions.
[217] La synthèse suivante de la gestion des risques dans le cadre de l'apprentissage automatique est tirée du document suivant: Future of Privacy Forum, *Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models*, 2018.

l'opportunité, la duplication, la validité, la disponibilité et l'origine des données d'entrée. Il peut également être important de mettre en place des mécanismes permettant de tester, d'actualiser et de contrôler le modèle au fil du temps.

Du côté des ***données de sortie***, divers processus peuvent être appliqués pour réduire le risque de voir les modèles d'apprentissage automatique aboutir à des résultats défavorables. Des mécanismes de détection des biais peuvent être mis en place pour garantir qu'aucun groupe de population n'est discriminé, ou du moins que les biais sont quantifiés et réduits au maximum. Il peut parfois être nécessaire de restreindre l'utilisation de certains types de données dans le modèle. Les données de sortie peuvent également être analysées pour détecter les proxys susceptibles de conduire à des discriminations, basées notamment sur le genre, l'appartenance ethnique ou encore l'origine géographique (code postal). Cette démarche suppose que des avocats renseignent les parties prenantes sur les types de caractéristiques qui constitueraient des fondements illégaux favorisant la discrimination. Une surveillance constante, grâce à une représentation statistique des données de sortie, devrait également améliorer la détection des anomalies, des boucles de rétroaction et d'autres comportements répréhensibles. Là encore, documenter ces processus et effectuer des tests en continu permettra d'améliorer et d'élargir la compréhension des risques présentés par un modèle donné.

L'évaluation des risques s'étend aux données d'entrée et de sortie, ainsi qu'à la création et au fonctionnement des algorithmes. L'institut de recherche AINow[218] a proposé que les organismes publics procèdent à des "évaluations de l'impact algorithmique", y compris lors de l'acquisition de données et de logiciels, et lors du fonctionnement des processus décisionnels automatisés, dans le

Principes FEAT de l'Autorité monétaire de Singapour

4. Les décisions axées sur le modèle AIDA sont régulièrement réexaminées afin de s'assurer que les modèles se comportent comme prévu.

5. L'utilisation du système AIDA est conforme aux normes déontologiques, aux valeurs et aux codes de conduite de l'entreprise.

6. Les décisions axées sur le modèle AIDA doivent respecter au moins les mêmes normes déontologiques que les décisions humaines.

Version préliminaire de *Digital Credit Standards* (normes relatives au crédit numérique) de la Smart Campaign

Indicateur 2.1.3.0

Si l'analyse de la capacité de remboursement est automatisée (par exemple, grâce à l'utilisation d'un algorithme), l'efficacité du système à prédire la capacité de remboursement des clients est évaluée par un service indépendant de l'équipe chargée de l'élaboration de l'algorithme au sein de l'organisation (par exemple, le service d'audit interne, la direction générale ou un autre département). Des recommandations sont ensuite formulées pour améliorer les résultats de l'algorithme, puis rapidement mises en œuvre.

Indicateur 2.1.10.0

Le fournisseur dispose d'un processus de contrôle interne rigoureux pour vérifier l'application uniforme des politiques et procédures relatives à la souscription de crédit, qu'il soit automatisé ou non (intervention du personnel).

Indicateur 2.1.10.1

Les critères guidant l'algorithme sont documentés, y compris les facteurs/types de variables utilisés et la justification du recours à ces facteurs. Un service indépendant au sein de l'organisation évalue régulièrement la cohérence et la conformité de la logique appliquée, de l'algorithme et de ses résultats. Il existe des preuves documentées des tests effectués et des mesures correctives qui ont été prises.

[218] https://ainowinstitute.org/.

cadre d'un vaste ensemble de mesures de responsabilisation[219].

Les responsables du traitement des données doivent définir conjointement les résultats escomptés ainsi que les résultats accidentels à éviter (en collaboration avec les équipes chargées des questions juridiques et de la conformité), et être prêts à corriger ou à retirer tout modèle problématique de son utilisation. Si les données de sortie risquent d'enfreindre la protection des usagers, la confidentialité des informations, les lois contre la discrimination ou autres, les entreprises doivent être prêtes à mettre en place une stratégie pour coopérer avec les autorités. Par exemple, les directives de l'État de Californie sur les permis relatifs aux véhicules autonomes contiennent des dispositions particulières sur la manière dont une entreprise doit interagir avec les forces de l'ordre en cas d'accident ou de tout autre imprévu.

Le bon fonctionnement des algorithmes, y compris la prévention des préjudices, passe en partie par une maintenance continue. Certains en appellent à une obligation légale permanente consistant à surveiller les résultats des algorithmes, à instaurer des mécanismes de retour d'informations (de dépôt de plaintes, par exemple), à procéder à des inspections et à corriger les modèles[220]. Des questions aussi complexes ne sont pas à la portée des usagers, qui manquent d'expertise et de ressources. Il importera parfois de favoriser une intervention humaine dans le processus de contrôle, non pas simplement dans le cadre d'un recours entrepris par un usager, mais dans le cadre du processus décisionnel à proprement parler. Une implication humaine de ce type doit être étudiée de manière réfléchie.

8.2 Intégrer la confidentialité des données à la conception

Pour aborder efficacement les questions de protection des consommateurs et de confidentialité des données dans le cadre des mégadonnées et de l'apprentissage automatique, il ne faudra pas se contenter du strict minimum en ce qui concerne le respect des lois en vigueur, mais aller au-delà de ce cadre réglementaire. Il s'agira notamment de concevoir des produits et des services dans l'optique de réduire au maximum les atteintes à la vie privée. Les sept principes liés à l'intégration du principe de confidentialité au processus de conception, élaborés sous la houlette d'Ann Cavoukian[221] sont les suivants:

1. Être proactif plutôt que réactif, et privilégier les mesures préventives par rapport aux mesures correctives, en anticipant et en empêchant tout événement portant atteinte à la vie privée.
2. Activer les paramètres de confidentialité par défaut, de sorte que les usagers n'aient à modifier aucun paramètre pour protéger leur vie privée (acceptation privilégiée par rapport au refus).
3. Intégrer le principe de respect de la vie privée dès la conception (notamment avec la fonction de portabilité des données), plutôt qu'après, afin qu'il fasse partie intégrante du système sans en compromettre la fonctionnalité.
4. Adopter une approche qui profite à toutes les parties, et qui bénéficie de la confiance accrue des usagers et d'un risque moindre de violation des données.

---

[219] Reisman, D., Schultz, J., Crawford, K. et Whittaker, M., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, avril 2018. Disponible à l'adresse suivante: https://ainowinstitute.org/aiareport2018.pdf.

[220] Initiative mondiale de l'IEEE (voir note de bas de page 223), p. 156.

[221] Cavoukian, A., *Privacy by Design: The Seven Foundational Principles* (Commissaire à l'information et à la protection de la vie privée de l'Ontario, 2011), disponible à l'adresse suivante: https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf; et Medine, D., *Privacy by Design for Financial Services*, disponible à l'adresse suivante: https://www.livemint.com/Opinion/1ShpKAOC59VlXiwgCkVv8O/Privacy-by-design-for-financial-services.html.

5. Garantir la sécurité de bout en bout, en veillant à la collecte, à la conservation et à la destruction sécurisées des données tout au long de leur cycle de vie (y compris au chiffrement du stockage et du transfert des données).
6. Faire preuve de visibilité et de transparence, en adoptant des politiques et en tenant des registres pour permettre un contrôle interne et une vérification indépendante.
7. Faire preuve de respect à l'égard de la vie privée des utilisateurs, en leur donnant accès aux données les concernant, ainsi que la possibilité de les contester, les corriger, les compléter et les mettre à jour.

Cette démarche nécessitera une ingénierie de la confidentialité, aussi bien dans le cadre de l'élaboration des produits que de la formation des informaticiens. Par exemple, l'Université Carnegie Mellon propose un master en technologie de l'information spécialisé dans l'ingénierie de la confidentialité qui aborde ce type de questions[222].

8.3 Éthique et autorégulation

Au-delà de la gestion et de l'ingénierie, des efforts plus larges sont actuellement déployés pour faire évoluer les attitudes sous-jacentes et la prise de conscience des personnes travaillant dans les technologies. Les efforts d'autorégulation s'appuient sur les principes proposés par les parties prenantes du secteur et d'autres acteurs. Ils se concentrent sur l'exactitude, l'équité, la responsabilité et la transparence, la croissance durable et le respect de la vie privée[223]. Il s'agit notamment des mesures prises par la communauté des ingénieurs pour instaurer une éthique de l'intelligence artificielle et de la prise de décisions autonome.

[222] Ledit programme comprend les thématiques suivantes: 1) concevoir des produits et des services de pointe qui s'appuient sur les mégadonnées tout en préservant la vie privée; 2) proposer et évaluer des solutions visant à atténuer les risques liés à la protection de la vie privée; 3) comprendre la manière dont les technologies d'amélioration de la confidentialité peuvent être utilisées afin de réduire les risques liés à la protection de la vie privée; 4) utiliser des techniques pour agréger et désidentifier les données, et comprendre les limites du processus de désidentification; 5) comprendre les cadres de réglementation et d'autorégulation en vigueur en matière de protection de la vie privée; 6) comprendre les problèmes actuels en matière de protection de la vie privée qui sont liés à la technologie; 7) évaluer les risques liés à la protection de la vie privée, contrôler la conformité, résoudre les incidents et intégrer le principe de protection de la vie privée aux différentes phases du cycle de vie de l'ingénierie logicielle; 8) effectuer une évaluation élémentaire de la facilité d'utilisation afin de connaître les fonctionnalités et les processus liés à la protection de la vie privée qui sont acceptés par les utilisateurs; et 9) mettre à disposition son expertise en matière de protection de la vie privée, en collaboration avec les équipes interdisciplinaires. Master en technologie de l'information spécialisé en ingénierie de la confidentialité. Voir également: https://bigid.com/the-advent-of-privacy-engineering/.

[223] Voir, par exemple: OCDE, "OECD Moves Forward on Developing Guidelines for Artificial Intelligence (AI)", 20 février 2019, disponible à l'adresse suivante: http://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm; et INSTITUTE OF BUSINESS ETHICS, BUSINESS ETHICS AND ARTIFICIAL INTELLIGENCE, 2018, p. 2-3.

Des organismes tels que l'Association for Computing Machinery (ACM)[224] et l'IEEE en sont des exemples[225], ainsi que le Partnership on AI[226], la Software & Information Industry Association (SIIA)[227], et des entreprises telles que Google[228] et Microsoft[229]. Leurs efforts s'accompagnent de travaux d'organisations telles que Fairness, Accountability, and Transparency in Machine Learning (FAT/ML)[230],

[224] Informatics Europe et le Comité européen de politique technologique de l'Association for Computing Machinery, "When computers decide: European Recommendations on Machine-Learned Automated Decision Making", 2018. Disponible à l'adresse suivante: http://www.informatics-europe.org/news/435-ethics_adm.html.

[225] IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, version 2, 2018. Disponible à l'adresse suivante: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf. Une première version a été publiée à des fins de consultation en 2016 (disponible à l'adresse suivante: http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf).

[226] Partnership on AI, *Tenets.* Disponible à l'adresse suivante: https://partnershiponai.org/about/#:~:text=Our%20Tenets,-We%20believe%20that&text=We%20will%20educate%20and%20listen,work%2C%20and%20address%20their%20questions.&text=We%20are%20committed%20to%20open,and%20legal%20implications%20of%20AI. .

[227] Software and Information Industry Association, *Ethical Principles for Artificial Intelligence and Data Analytics*, 15 septembre 2017. Disponible à l'adresse suivante: http://www.siia.net/Portals/0/pdf/Policy/Ethical%20Principles%20for%20Artificial%20Intelligence%20and%20Data%20Analytics%20SIIA%20Issue%20Brief.pdf?ver=2017-11-06-160346-990.

[228] Google, "AI at Google: our principles", 7 juin 2018. Disponible à l'adresse suivante: https://www.blog.google/technology/ai/ai-principles/. Voir aussi: Google, PERSPECTIVES ON ISSUES IN AI GOVERNANCE. Disponible à l'adresse suivante: https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf.

[229] Microsoft, *Our approach*. Disponible à l'adresse suivante: https://www.microsoft.com/en-us/ai/our-approach-to-ai.

[230] Fairness, Accountability, and Transparency in Machine Learning, *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.* Disponible à l'adresse suivante: https://www.fatml.org/resources/principles-for-accountable-algorithms.

Privacy International[231], le Future of Life Institute[232], le Center for Democracy & Technology (CDT)[233] et la Leadership Conference[234].

Dans le domaine des services financiers, comme mentionné dans l'introduction (section 3), la Smart Campaign a permis d'établir des indicateurs provisoires sur les algorithmes et les décisions automatisées fondées sur les données dans le cadre de ses normes de crédit numérique (voir l'annexe B "Normes de la Smart Campaign relatives au crédit numérique"), dont bon nombre ont été cités dans ce rapport. La Smart Campaign[235] est hébergée par le Center for Financial Inclusion d'Accion[236]. Son objectif est d'élaborer et de promouvoir des normes d'autorégulation pour la protection des usagers (et autres clients) dans le domaine de l'inclusion financière. Elle assure notamment la gestion d'un programme de certification à l'intention des fournisseurs de services financiers. La Smart Campaign et MFR[237], une agence de notation indépendante qui attribue une grande partie des certifications de la Smart Campaign relatives à la protection des clients, ont préparé des normes en la matière à l'intention des fournisseurs de crédit numérique. Celles-ci ont été testées auprès de deux fournisseurs de services financiers qui opèrent au Kenya et s'appuient sur des interactions automatisées avec les usagers (4G Capital[238] et Tala[239]). Une version révisée a été publiée par la suite. Le document de la Banque mondiale intitulé *Use of Alternative Data to Enhance Credit Reporting to Enable Access to Digital Financial Services by Individuals and*

---

[231] Privacy International, PRIVACY AND FREEDOM OF EXPRESSION IN THE AGE OF ARTIFICIAL INTELLIGENCE, 2018. Disponible à l'adresse suivante: https://privacyinternational.org/sites/default/files/2018-04/Privacy%20and%20Freedom%20of%20Expression%20%20In%20the%20Age%20of%20Artificial%20Intelligence.pdf.

[232] Future of Life Institute, *Asilomar AI Principles*, 2017. Disponible à l'adresse suivante: https://futureoflife.org/ai-principles/. Composé d'universitaires, de dirigeants du secteur de l'intelligence artificielle et d'autres acteurs divers, ce groupe a formulé 13 principes: **1) Sécurité:** les systèmes d'intelligence artificielle doivent être sûrs et sécurisés tout au long de leur durée de vie opérationnelle, et de manière vérifiable lorsque ce s'avère possible et réaliste; 2) **Transparence face aux défaillances:** si un système d'intelligence artificielle entraîne un préjudice, il doit pouvoir être possible d'en déterminer les raisons; 3) **Transparence judiciaire:** toute implication d'un système autonome dans une prise de décision judiciaire doit permettre des explications satisfaisantes et vérifiables de la part d'une autorité humaine compétente; 4) **Responsabilité:** les concepteurs et les créateurs de systèmes avancés d'intelligence artificielle sont parties prenantes quant aux implications morales de leur utilisation (à bon ou à mauvais escient) et de leurs actes, et ont la possibilité et la responsabilité de moduler ces implications; 5) **Alignement des valeurs:** les systèmes d'intelligence artificielle hautement autonomes doivent être conçus de manière à ce que leurs objectifs et leurs comportements soient alignés sur les valeurs humaines tout au long de leur fonctionnement; 6) **Valeurs humaines:** les systèmes d'intelligence artificielle doivent être conçus et exploités de manière à être compatibles avec les idéaux de respect de la dignité humaine, des droits, des libertés et de la diversité culturelle; 7) **Confidentialité des données personnelles**: les personnes doivent avoir le droit d'accéder aux données personnelles qu'elles ont communiquées, et d'assurer leur gestion et leur contrôle, étant donné que les systèmes d'intelligence artificielle ont la capacité d'analyser et d'utiliser ces données; 8) **Liberté et respect de la vie privée:** l'application de l'intelligence artificielle aux données personnelles ne doit pas restreindre de manière déraisonnable la liberté réelle ou perçue des personnes; 9) **Avantages partagés:** les technologies reposant sur l'intelligence artificielle doivent bénéficier au plus grand nombre de personnes possible et leur donner les moyens d'agir; 10) **Prospérité partagée:** la prospérité économique permise par l'intelligence artificielle doit être partagée de façon à bénéficier à l'ensemble de l'humanité; 11) **Contrôle humain:**. les acteurs humains doivent choisir si et comment déléguer des décisions à des systèmes d'intelligence artificielle, afin d'atteindre des objectifs fixés par les humains; 12) **Non-subversion:** le pouvoir conféré par le contrôle de systèmes d'intelligence artificielle hautement avancés doit respecter et améliorer, plutôt que bouleverser, les processus sociaux et civiques dont dépend le dynamisme de la société; 13) **Course aux armements de l'intelligence artificielle:**. il faut éviter une course aux armements dans le domaine des armes autonomes létales.

[233] Center for Democracy & Technology, *Digital Decisions.* Disponible à l'adresse suivante: https://cdt.org/insights/digital-decisions-tool/

[234] The Leadership Conference on Civil and Human Rights, *Civil Rights Principles for the Era of Big Data*, 27 février 2014. Disponible à l'adresse suivante: https://civilrights.org/civil-rights-principles-era-big-data/.

[235] http://www.smartcampaign.org/.

[236] https://www.centerforfinancialinclusion.org/.

[237] https://www.mf-rating.com/fr/.

[238]

[239] https://tala.co/.

SMEs operating in the Informal Economy (voir section 3 et note de bas de page 14) est un autre exemple représentatif des orientations destinées aux fournisseurs de services financiers.

De telles mesures ne suffisent pas à garantir l'équité, la responsabilité et la transparence, mais elles fournissent un lexique et un système de valeurs qui permettent une communication beaucoup plus rapide sur ces questions et facilitent le développement de la gestion des risques, de l'ingénierie et d'autres mesures nécessaires à une meilleure protection de la vie privée des usagers.

9 Domaines à approfondir

Le présent document étudie les lacunes légales et réglementaires en matière de protection des usagers et de confidentialité des données en ce qui concerne les techniques liées aux mégadonnées et à l'apprentissage automatique, en particulier lorsqu'il s'agit de prendre des décisions relatives aux services proposés aux usagers. Les exigences traditionnelles consistant à renseigner l'usager sur la finalité de l'utilisation des données à caractère personnel alors que celle-ci est encore incertaine, ou à obtenir son consentement à l'égard de quelque chose qui dépasse de loin son entendement, sont remises en question. Les risques liés à l'inexactitude des données saisies ou au traitement partial et discriminatoire dans les processus décisionnels automatisés soulèvent également plusieurs questions complexes, par exemple: comment garantir l'absence de tout traitement injuste des usagers? Assurer la transparence des décisions générées par les algorithmes ou mettre en évidence les préjudices qui ont été directement provoqués par les technologies de l'intelligence artificielle représente également un défi dans le cadre de la législation et de la réglementation en matière de protection des usagers et de confidentialité des données.

Il existe plusieurs domaines pour lesquels il serait utile de poursuivre ces travaux afin d'élaborer des normes applicables aux mégadonnées et à l'apprentissage automatique, en vue de trouver un équilibre entre la liberté d'innover et la protection des usagers et de leurs données, en particulier concernant les questions suivantes:

1. ***Améliorer la pertinence du consentement à l'utilisation et au partage des données personnelles.*** Il s'agirait notamment d'améliorer la transparence et la clarté des informations communiquées aux usagers sur l'utilisation qui peut être faite de leurs données, notamment en fournissant des explications intelligibles. Une réglementation plus stricte des conditions de consentement peut également venir compléter la mise sur le marché des technologies en la matière. Lorsque l'utilisation de données à caractère personnel dépasse le strict cadre du service direct à l'usager pour impliquer des tiers, le fait de donner à l'usager les moyens de porter un jugement pertinent et éclairé sur la communication de ses données personnelles à de tierces parties pourra être envisagé.

2. Envisager de ***recadrer la réglementation relative à l'utilisation et au partage des données personnelles*** lorsqu'il est tout simplement irréaliste d'attendre des usagers qu'ils comprennent ce qu'une circulation généralisée de leurs données personnelles implique pour eux**.** Il peut s'agir de ne pas se contenter d'obtenir le consentement de l'usager sur des questions qui dépassent son entendement, mais de veiller à ce que celui-ci soit mieux informé et que les transferts de données le concernant soient mieux contrôlés, tout en le protégeant contre toute utilisation de données personnelles à laquelle il ne pourrait raisonnablement pas s'attendre.

3. Élaborer des normes pour l'intégration des ***principes de confidentialité à la conception*** de modèles d'intelligence artificielle et d'apprentissage automatique. Dans la lignée des principes

élaborés par Ann Cavoukian (voir section 8.2), il pourrait s'agir de normes concernant 1) une approche proactive de la conception, 2) l'utilisation de paramètres par défaut en faveur de la protection de la vie privée, 3) l'adoption du principe de confidentialité dès la conception, 4) la confiance des usagers en tant qu'axe prioritaire, 5) la sécurité de bout en bout, 6) l'accès des usagers aux informations et la possibilité de contester, de corriger, de compléter et de mettre à jour les données les concernant, et 7) la création, l'enregistrement et la communication de journaux d'événements et de pistes de vérification du processus de conception, afin de permettre des révisions et de garantir que ces journaux et pistes sont codés dans le système.

4. Élaborer des ***normes éthiques pour la programmation informatique de l'intelligence artificielle*** auxquelles la communauté des développeurs peut se rapporter pour aborder les différents types de questions examinées dans le présent document. Ces normes pourront en outre poser les jalons d'une discussion continue permettant de soulever et d'aborder de nouvelles questions.

5. Élaborer des ***normes relatives à une analyse déductive acceptable.*** Celles-ci pourraient porter sur l'évaluation des données de sortie et des décisions des modèles d'apprentissage automatique au regard des principes de protection de la vie privée et de lutte contre la discrimination. Elles pourraient également aborder la question de savoir si les déductions opérées sur le plan personnel (opinions politiques, orientation sexuelle, santé, etc.) à partir de différentes sources de données (navigation sur Internet, notamment) sont acceptables ou portent atteinte à la vie privée, en fonction du contexte. Il pourrait également s'agir d'élaborer des normes relatives à la fiabilité des déductions, en particulier celles qui sont importantes sur le plan social, qui présentent un risque ou qui ont des répercussions juridiques significatives, et qui concernent les groupes protégés. Des normes pourraient en outre être élaborées pour vérifier les déductions avant et après le déploiement des données. Cette démarche pourra nécessiter des approches différentes selon les types de services proposés.

6. Élaborer des ***normes relatives à l'explication des décisions automatisées***, notamment en défendant la pertinence des données utilisées par le système pour effectuer des déductions, la pertinence de ces déductions au regard de chaque type de décision automatisée, ainsi que l'exactitude et la fiabilité statistique des données et des méthodes utilisées. Il pourrait s'agir d'encourager les développeurs de modèles de notation à communiquer aux usagers (et, si nécessaire, aux organismes de réglementation) les principales caractéristiques utilisées dans chaque modèle, ainsi que leur pondération relative, et de veiller à fournir la documentation et les pistes de vérification requises en cas de procédure judiciaire. Il serait également possible d'envisager l'utilisation d'éléments contrefactuels pour informer l'usager de la manière dont le système automatisé pourrait, avec des caractéristiques d'entrée différentes, aboutir à des décisions différentes. Lorsque les explications contrefactuelles sont considérées comme valides, leur communication en aval des décisions pourrait faire l'objet de normes.

7. Mettre en place de ***bonnes pratiques dans les processus i) permettant aux usagers de bénéficier d'une intervention humaine***, et ii) facilitant la détermination du degré d'intervention humaine nécessaire pour maintenir l'intégrité et la valeur du modèle, tout en permettant un dialogue entre l'usager et un autre humain.

8. Élaborer les ***principes relatifs aux bonnes pratiques internationales et harmoniser les mécanismes de responsabilisation***, y compris les procédures de contestation des décisions automatisées, les normes visant à démontrer tout préjudice *prima facie* et, en définitive, les cadres pour l'évaluation de la responsabilité associée aux modèles d'intelligence artificielle et d'apprentissage automatique en matière de conception et d'exploitation.

Annexe A (Principes FEAT de l'Autorité monétaire de Singapour)

Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector[240].

Équité

Justification

1. Les individus ou groupes d'individus ne sont pas systématiquement désavantagés par les décisions prises grâce aux technologies d'AIDA, sauf si ces décisions peuvent être justifiées.
2. L'utilisation d'attributs personnels comme facteurs d'entrée pour prendre des décisions à partir du modèle AIDA est justifiée.

Exactitude et biais

3. Les données et les modèles utilisés pour prendre des décisions reposant sur les technologies d'AIDA sont régulièrement examinés et validés à des fins d'exactitude et de pertinence, et pour limiter les biais involontaires.
4. Les décisions axées sur le modèle AIDA sont régulièrement réexaminées afin de s'assurer que les modèles se comportent comme prévu.

Déontologie

5. L'utilisation du système AIDA est conforme aux normes déontologiques, aux valeurs et aux codes de conduite de l'entreprise.
6. Les décisions axées sur le modèle AIDA doivent respecter au moins les mêmes normes déontologiques que les décisions humaines.

Responsabilité

Responsabilité interne

7. L'utilisation des technologies d'AIDA pour prendre des décisions est approuvée par une autorité interne appropriée.
8. Les entreprises qui utilisent les technologies d'AIDA sont responsables des modèles d'AIDA élaborés en interne ou provenant de l'extérieur.
9. Les entreprises qui utilisent les technologies d'AIDA sensibilisent de manière proactive la direction et le conseil d'administration à leur utilisation.

[240] Disponible à l'adresse suivante: https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf.

Responsabilité externe

10. Les personnes disposent de canaux leur permettant de se renseigner sur les décisions les concernant prises grâce aux technologies d'AIDA, de faire appel et de demander leur révision.
11. Les données supplémentaires communiquées par les personnes concernées, à condition qu'elles soient pertinentes et vérifiées, sont prises en compte au moment de revoir les décisions fondées sur le modèle AIDA.

Transparence

12. Pour accroître la confiance du public, les personnes concernées sont informées de manière proactive de l'utilisation des technologies d'AIDA dans le cadre d'une communication générale.
13. Les personnes concernées qui en font la demande reçoivent des explications claires sur les données utilisées pour prendre des décisions les concernant grâce aux technologies d'AIDA et sur la manière dont ces données influent sur lesdites décisions.
14. Les personnes concernées qui en font la demande reçoivent des explications claires sur les conséquences que les décisions prises grâce aux technologies d'AIDA peuvent avoir pour elles.

Annexe B (Normes de la Smart Campaign relatives au crédit numérique)

Indicateurs provisoires sur les algorithmes et les décisions automatisées fondées sur des données

Principe relatif à la protection des clients nº 2: Prévention du surendettement

Indicateur 2.1.3.0
Si l'analyse de la capacité de remboursement est automatisée (par exemple, grâce à l'utilisation d'un algorithme), l'efficacité du système à prédire la capacité de remboursement des clients est évaluée par un service indépendant de l'équipe chargée de l'élaboration de l'algorithme au sein de l'organisation (par exemple, le service d'audit interne, la direction générale ou un autre département). Des recommandations sont ensuite formulées pour améliorer les résultats de l'algorithme, puis rapidement mises en œuvre.

Indicateur 2.1.5.0
Les données et les analyses de souscription sont actualisées à chaque cycle de prêt afin de suivre l'évolution de la situation du client.

Indicateur 2.1.10.0
Le fournisseur dispose d'un processus de contrôle interne rigoureux pour vérifier l'application uniforme des politiques et procédures relatives à la souscription de crédit, qu'il soit automatisé ou non (intervention du personnel).

Indicateur 2.1.10.1
Les critères guidant l'algorithme sont documentés, y compris les facteurs/types de variables utilisés et la justification du recours à ces facteurs. Un service indépendant au sein de l'organisation évalue régulièrement la cohérence et la conformité de la logique appliquée, de l'algorithme et de ses résultats. Il existe des preuves documentées des tests effectués et des mesures correctives qui ont été prises.

Principe relatif à la protection des clients nº 5: Traitement équitable et respectueux

Indicateur 5.2.1.0
En fonction de son origine ethnique, son genre, son âge, sa situation au regard du handicap, son affiliation politique, son orientation sexuelle, sa caste et sa religion, une personne appartiendra à une catégorie protégée ou non.

Indicateur 5.2.3.0
Les algorithmes sont conçus pour réduire le risque de discrimination des usagers liée aux catégories protégées.

Indicateur 5.2.3.1
Après une phase d'apprentissage initiale, le fournisseur effectue une analyse des liens entre les variables non discriminatoires et les variables discriminatoires afin de vérifier l'absence de biais involontaire dans les décisions de crédit automatisées.

Indicateur 5.2.3.2
Si le fournisseur confie le développement de l'algorithme à une tierce partie, il doit exiger de cette dernière qu'elle respecte les normes exposées à l'indicateur ci-dessus. La tierce partie communique au fournisseur les informations suivantes: paramètres et documentation de l'algorithme, supports de formation fournis à l'équipe et documents relatifs à l'historique des tests antérieurs (comprenant

notamment la date, une description et le résultat de chaque test, les éléments de discrimination identifiés et les mesures correctives prises, le cas échéant).

Principe relatif à la protection des clients nº 6: Confidentialité, sécurité et intégrité des données

Indicateur 6.1.1.0
Des politiques et des processus sont en place et tenus à jour pour préserver la confidentialité, la sécurité et l'exactitude des informations personnelles, transactionnelles et financières des clients. Ceux-ci concernent la collecte, l'utilisation, la diffusion et la conservation des données.

Indicateur 6.1.1.1
Le fournisseur a évalué et consigné les informations personnelles dont il a besoin de la part de ses clients afin d'assurer le service proposé (par exemple, identité, transactions, etc.). La collecte, le partage et la durée de conservation des données personnelles sont réduits au strict nécessaire et explicitement justifiés par les opérations requises pour la prestation des services ou par le cadre réglementaire. L'évaluation a permis d'identifier les risques que la collecte, le traitement, la conservation et la communication des données personnelles représentent pour la vie privée des usagers.

Indicateur 6.1.1.6
Les données à caractère personnel doivent être i) pertinentes au regard des finalités pour lesquelles elles seront utilisées et, dans la mesure nécessaire à ces fins, ii) exactes, complètes et à jour.

Indicateur 6.2.1.0
Les personnes concernées sont invitées à consentir aux utilisations qui seront faites de leurs données personnelles. Les demandes de consentement expliquent clairement, dans un langage simple et dans la langue locale, la manière dont les données seront utilisées. Un consentement distinct est requis pour: a) le partage des données avec certaines parties tierces (à identifier clairement) dans le cadre de la prestation de services; b) la communication des données aux agences d'évaluation du crédit; c) l'utilisation des données à des fins de marketing; d) la vente de données à des tiers; et e) l'utilisation des données de géolocalisation. S'agissant des services reposant sur des données de service complémentaire non structurées (USSD) ou sur des messages textes, fournir les hyperliens vers les accords de divulgation ne suffit pas.

Indicateur 6.2.2.0
Le droit de renoncer à un service et de retirer l'autorisation accordée à une organisation d'utiliser des données (de quelque type que ce soit) doit être clairement indiqué et accessible aux usagers, de même que les conséquences d'un tel retrait.

Indicateur 6.2.3.0
Les usagers ont le droit d'obtenir du fournisseur la confirmation que celui-ci possède ou non des données les concernant, et si cette demande est rejetée, ils ont le droit d'en connaître les raisons.

Indicateur 6.2.3.1
Les usagers ont le droit de se voir communiquer les données les concernant dans des délais raisonnables, sans frais excessifs et dans des termes compréhensibles.

Indicateur 6.2.3.2
Les consommateurs ont le droit de contester les données les concernant et, si ladite contestation aboutit, de faire effacer, rectifier, compléter ou modifier ces données.