



AGORA

ALGORITHMES: GARDER LE CONTRÔLE

Les systèmes algorithmiques s'imposent dans les sphères publique et privée avec leur lot de défis, de la conception à l'usage: erreurs, biais, discriminations... Comment préserver la transparence et le libre arbitre humain ?

En janvier 2021, le Premier ministre néerlandais ainsi que son gouvernement sont contraints de démissionner. En cause? Un système algorithmique biaisé. Mis en place pour détecter les demandes d'allocations familiales erronées et potentiellement frauduleuses, il a accusé à tort des dizaines de milliers de personnes de fraude. De plus, il ciblait surtout les personnes n'étant pas de nationalité néerlandaise, car il considérait que cela constituait un facteur de risque. Au Royaume-Uni, à l'été 2020, un algorithme de classement scolaire provoquait un tollé national. Conséquence de son utilisation, 40 % des étudiants ont obtenu une notation inférieure ne reflétant pas leurs résultats antérieurs. Des lycéens scolarisés dans des zones défavorisées ne pouvaient alors plus accéder aux universités prestigieuses pour lesquelles ils étaient pressentis. Aux États-Unis, Compas, l'algorithme utilisé pour guider la détermination de la peine en prédisant la probabilité d'une récidive criminelle est accusé, dès 2016, de pénaliser les Afro-Américains. Alors que le recours à des systèmes algorithmiques s'intensifie dans la sphère publique (demande d'allocations, candidature d'entrée à l'université ou à un emploi, obtention d'un prêt, souscription d'une assurance, etc.), les scandales se multiplient et la défiance s'installe. Dans son livre *Algorithmes : la bombe à retardement*, paru en 2017, la mathématicienne Cathy O'Neil les qualifie même d'armes de destruction mathématique. Les problèmes posés par les algorithmes sont nombreux, mais pour les professionnels, les questions d'opacité, des biais et de l'adaptation du droit à ces nouveaux outils prédominent. Le premier grief formulé à l'encontre des algorithmes d'apprentissage automatique (*machine learning*) prenant leur décision à partir d'un historique de données repose sur leur complexité. «Quand il s'agit de modèles statistiques élémentaires où seuls quelques facteurs ou variables interviennent, la décision qui en découle est interprétable ou compréhensible. Mais lorsqu'il s'agit d'un modèle ou d'un algorithme où interviennent des milliers, voire des millions de paramètres pour

«Un algorithme simple, explicable, ne permet pas d'approcher le réel avec assez de qualité et de précision. Parfois, ce n'est pas tant l'algorithme qui est complexe, difficile à interpréter, que la réalité sous-jacente.»



Philippe BESSE
Professeur émérite
à l'université de Toulouse,
INSA, pôle génie
mathématique
et modélisation

26

construire une décision, cette dernière devient totalement opaque. De toute façon, si vous faites face à une réalité complexe et non linéaire, un algorithme simple ne peut en rendre compte. Le vivant est un exemple évident: il y a des milliers de facteurs, génétiques ou environnementaux et leurs interactions, qui peuvent influer sur le développement de certaines pathologies. Un algorithme simple, explicable, ne permet pas d'approcher le réel avec assez de qualité et de précision. Parfois, ce n'est pas tant l'algorithme qui est complexe, difficile à interpréter, que la réalité sous-jacente», estime Philippe Besse, professeur émérite à l'université de Toulouse (INSA, département de génie mathématique et modélisation). «Si les algorithmes de machine learning ont l'avantage de pouvoir gérer de gros volumes de données hétérogènes et potentiellement non structurées, ils ont l'inconvénient d'introduire un effet boîte noire. Ce compromis performance-transparence peut faire apparaître un certain nombre de questions auxquelles il convient de répondre au risque de freiner dans certains cas le déploiement de ce type de modèle. Qu'est-ce qu'il fait? Est-ce que je suis serein avec ce résultat? C'est un problème majeur», renchérit Marc Juillard, actuair certifié IA, directeur data lab et programme IA chez Société générale Insurance.

Lutter contre les biais et les discriminations

La capacité du *machine learning* à reproduire des biais et des discriminations pose également question. «Le modèle crée rarement un biais. Cependant, il reproduit, voire amplifie, ceux présents dans les données d'entrée. Un exemple bien connu est celui du moteur de recrutement d'Amazon. Il y a quelques années, Amazon a souhaité mettre en place un algorithme de scoring des CV afin de prioriser les candidats à rencontrer. Amazon a donc naturellement utilisé son historique de CV et d'embauches afin d'entraîner son algorithme. Problème: le modèle ainsi construit présentait, certes, une très bonne performance, mais il appliquait une discrimination vis-à-vis des femmes. Amazon a pu en détecter la principale raison: par le passé, elle engageait principalement des hommes, car les profils féminins étaient à l'époque plus rares dans la tech. Le modèle a détecté ce phénomène et l'a relié non pas au sexe du candidat, cette information étant absente, mais aux hobbies qui étaient indirectement genrés. Il a ensuite intégré dans son scoring cet impact hobbies, pour aboutir in fine à un résultat extrêmement sexiste», explique Marc Juillard. L'algorithme n'est donc pas le seul élément de l'équation qui doit être contrôlé. «La problématique de la qualité des données, mais aussi celle de la capacité à assembler des données qui soient pertinentes et autorisées, ne doivent pas être négligées. La surveillance s'imbrique en différents endroits, aux croisements entre les données, les modèles et les risques. C'est pour cela qu'il n'est pas évident de donner des règles simples de gouvernance», indique Christian Robert, professeur de statistiques et de science actuarielle à l'Ensaee.

Une situation qui réaffirme l'importance de l'esprit critique du modélisateur. Pour Arthur Charpentier, professeur au département mathématiques de l'Université du Québec à Montréal et actuair agrégé IA, «le doute scientifique revient avec la question de la fiabilité des données». «Les jeux de données d'apprentissage doivent être soigneusement constitués pour être statistiquement représentatifs, prolonge Philippe Besse. La sous-représentation de certains groupes peut par exemple entraîner une discrimination en termes d'erreur de prévision. C'est typiquement ce qui a été dénoncé pour la reconnaissance faciale. À savoir que le

taux d'erreur pour les femmes noires était de l'ordre de 30 %, alors que celui pour les hommes d'origine caucasienne était très faible. » Mais la découverte de la discrimination implique de lever un paradoxe lié aux données sensibles. « Pour savoir s'il y a un biais selon l'origine ethnique des personnes, il faut disposer de cette information sur l'origine ethnique et cela nécessite l'enregistrement de données sensibles, détaille Philippe Besse. Or, cet enregistrement est en principe interdit par le RGPD, sauf exceptions. Cette possibilité de prendre en compte les données sensibles devrait être renforcée par le projet de règlement européen : l'AI Act. Il y a en effet une forme d'hypocrisie que d'affirmer ne pas stocker ou prendre en compte l'origine ethnique pour éviter la discrimination. Cette information est très généralement présente dans d'autres variables (prénom, origine, adresse) susceptibles de produire une discrimination algorithmique qui ne pourra pas être détectée par un contrôle a posteriori sans la connaissance effective de l'origine. » « Masquer un problème le résout rarement, même si c'est la solution qui a longtemps été préconisée pour lutter contre les discriminations, appuie Arthur Charpentier. Le RGPD est venu rajouter une couche de complexité. Avec les données protégées, beaucoup d'assureurs se sont dit : "N'ayant plus le droit d'utiliser ces données, ne les gardons pas dans notre base." Observer des risques et savoir ce qui se passe, c'est important. Et je pense qu'il faut clairement distinguer ce qui sert à faire des études pour comprendre la sinistralité, et l'utilisation qui en est faite dans les prix et les traitements des assurés. »

La quête de l'explicabilité

L'opacité des algorithmes n'est pas une fatalité. La recherche s'intéresse ainsi à la question depuis une vingtaine d'années et des solutions techniques ont fait leur apparition. « Une batterie de surmodèles existe, Lime ou Shap pour ne citer

« Il faut clairement distinguer ce qui sert à faire des études pour comprendre la sinistralité, et l'utilisation qui en est faite dans les prix et les traitements des assurés. »



Arthur CHARPENTIER
Professeur au département mathématiques de l'Université du Québec à Montréal, actuaire agrégé IA

**DIGITAL
iNSURE**

**LA PRÉVOYANCE
RÉSOLUMENT TECH**



#DISTRIBUTION #GESTIONDÉLÉGUÉE
#MARKETPLACE #API #RISQUESPECIAUX

**CONFIEZ VOTRE
TRANSFORMATION DIGITALE
À LA FINTECH LEADER EN
ASSURANCE EMPRUNTEUR &
PRÉVOYANCE**

**Approche
Test & Learn**

**Algorithmes
d'aide à la
décision**

**Data
analyse**

www.digital-insure.fr



La
FRENCH TECH
FT120

« Il ne faut pas être pessimiste en se disant que la transparence entrave nécessairement la prédiction. »



Marc JUILLARD
Directeur data lab et programme IA chez Société générale Insurance, actuaire certifié IA

que les plus connus. Ces modèles complémentaires vont s'appuyer sur les données, ainsi que les prédictions pour nous fournir des éléments d'explicabilité, avance Marc Juillard. Le surmodèle vient synthétiser le message qu'a appris l'algorithme de machine learning. C'est un interprète. Et la question qui va être posée, c'est à quel point il doit être fin. Finalement, si un modèle d'explicabilité est extrêmement fin et fidèle à l'algorithme de machine learning initial, il va devenir aussi compliqué que l'algorithme initial. En d'autres termes, je n'aurais peut-être plus la barrière de la langue, mais j'aurais encore celle du vocabulaire.» Philippe Besse pointe que les risques algorithmiques (confidentialité, erreurs, opacité, discrimination) sont interdépendants. Ainsi, la recherche d'un modèle explicable impacte la qualité de prédiction et nuit donc à l'aide à la décision : « Si vous simplifiez un algorithme, vous perdez en précision. Il s'agit alors de rechercher un meilleur compromis entre interprétabilité et précision. » « Ce n'est pas parce que j'utilise du machine learning que je vais avoir in fine un tarif incompréhensible, évalue Marc Juillard. Je peux toujours réussir à expliquer mon tarif quitte à simplifier mon modèle. L'actuaire peut très bien réaliser son analyse statistique grâce à un algorithme de machine learning, intégrer les signaux détectés dans un autre modèle transparent de type GLM ou GAM. Ces démarches garantissent la transparence pour la tarification. Il ne faut pas être pessimiste en se disant que la transparence entrave nécessairement la prédiction. Nous pouvons énormément améliorer la connaissance du risque avec le machine learning tout en restant transparents. »

Mais l'explicabilité n'est pas qu'une affaire de technicien. « Pour renforcer la transparence des algorithmes, nous n'avons pas travaillé uniquement sur le décodage technique, mais aussi sur un décodage compréhensible pour des profils non techniques. Si le data scientist brasse de la donnée, il y a toujours un commanditaire, un responsable business et ils doivent être capables de parler le même langage. Nous avons pour cela créé Shapash, un outil permettant d'expliquer les conseils de nos modèles avec une interface compréhensible pour les concepteurs et les experts métier de la relation client, expose Olivier Baes, Business Owner IA et Data opérationnelle chez MAIF. Un outil qui permet également de renforcer la traçabilité des algorithmes. Grâce à Shapash, nous réalisons des rapports d'audit. Nous mettons à jour le livret de famille de l'algorithme avec son historique et son évolution. Nous documentons les échanges avec les utilisateurs. Un algorithme évolue avec le temps, car les données changent. Pour expliquer le passé, il faut connaître le pedigree de l'algorithme. Cela nous permettra sûrement demain d'avoir une interface d'échange avec les auditeurs. »

Placer l'humain au cœur du système

L'adoption d'un système algorithmique par des utilisateurs professionnels dépend de la manière dont ils sont associés à sa conception et formés. « Typiquement, un outil intéressant a été abandonné ou mis en retrait: Watson, développé par IBM », juge Sonia Desmoulin-Canselier, coauteure de *Décider avec les algorithmes ; quelle place pour l'Homme, quelle place pour le droit ?*, docteure en droit privé et chargée de recherche au CNRS. « Celui-ci était utilisé en médecine notamment. Son intérêt était qu'il ne se contentait pas de faire du traitement d'imagerie médicale : il mettait en relation des données variées issues d'éléments du dossier patient, d'imagerie, de la littérature scientifique, etc. Son entraînement avait été fait en collaboration avec des services hospitaliers spécialisés.

Donc c'était, théoriquement, un modèle de réussite. Mais lorsqu'il a été utilisé dans d'autres services, pays et contextes médicaux, les médecins ont constaté que ses propositions ne correspondaient pas à leurs habitudes et aux besoins de leur pratique. S'ils avaient été mieux informés de la manière dont l'outil a été entraîné et comment l'adapter à leurs besoins et à leur milieu de travail, les choses auraient sûrement été différentes. Je pense que l'effort doit véritablement porter sur la façon dont les professionnels utilisateurs sont formés, informés, et la manière dont nous mettons en place ces outils, non pas pour gagner tout de suite du temps et de l'argent, mais pour que le service rendu, public ou privé, soit meilleur. » « Tous nos concepteurs ont eu une sensibilisation aux questions éthiques posées par les algorithmes, souligne Olivier Baes. Cela s'inscrit dans notre politique de technologie au service de l'humain. Nous avons créé des guides de conduite inspirés du Ethics guidelines for trustworthy AI. Il faut placer l'humain au cœur du système. Par exemple, le conseiller qui bénéficie de la recommandation de l'algorithme doit posséder des clés de lecture de la décision afin de juger si la recommandation est pertinente dans le contexte. Il doit disposer de son libre arbitre et ne pas être enfermé par la décision de l'algorithme. »

La supposée neutralité de l'outil est également une illusion à dissiper. Pour Caroline Lequesne Roth, maître de conférences en droit public et directrice du Master Algorithmic Law & Data Governance de l'université Côte d'Azur, « l'affaire » Datajust démontre que cette idée demeure dans l'esprit des législateurs. Après avoir été saisi par la Quadrature du Net et d'autres associations, le 31 décembre 2021, le Conseil d'État a validé l'usage du système algorithmique Datajust—destiné à élaborer un référentiel indicatif d'indemnisation des préjudices corporels et servir d'aide à la décision dans le cadre de procédures de règlement amiable des litiges. « Cette négation de la normativité numérique en France est assez frappante, relève Caroline Lequesne Roth. Il n'est pas consi-

« Je pense que l'effort doit véritablement porter sur la façon dont les professionnels utilisateurs sont formés. »



Sonia DESMOULIN-CANSELIER
Coauteure de Décider avec les algorithmes



Guider.



Épauler.



Soutenir.

Nous sommes là pour permettre au plus grand nombre de construire leur vie en confiance.

déré que les systèmes sont susceptibles d'entraver ou de surajouter au droit. Pour le Conseil d'État, il s'agit simplement d'un équipement de l'administration, qui ne produit pas d'effet. La contrainte managériale qui s'exerce, le contexte de la décision, le fait que le juge va être plus ou moins dépendant, lié ou confiant dans les recommandations formulées par le système, ne sont pas pris en compte. » D'où l'importance d'adapter le droit à la formation des professionnels, selon Sonia Desmoulin-Canselier : « Actuellement, nous sommes peut-être dans la facilité, celle d'ajouter un article à la loi de bioéthique, par exemple. Parce que, effectivement, il est possible de le faire vite, personne ne va s'opposer à un nouveau droit à l'information du patient. Vous pouvez toujours attribuer des droits individuels extraordinaires, si le professionnel n'est pas en mesure de bien travailler, vos droits ne peuvent pas être correctement respectés. La législation, ce ne sont pas seulement des droits pour les justifiables. Il faut considérer qu'il y a un intérêt collectif à organiser les choses pour que chacun soit protégé », souligne-t-elle.

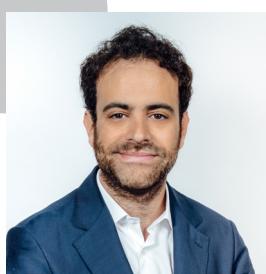
Réguler les systèmes algorithmiques

« Les algorithmes vont devenir si importants dans notre monde qu'il serait utile qu'un cadre réglementaire harmonisé aborde le sujet dans sa généralité, c'est-à-dire dans tous les domaines où ils interviennent, et non plus dans des domaines spécifiques, estime Benjamin Znaty, avocat au sein du département IP/IT du cabinet Taylor Wessing. Aux États-Unis, nous discutons depuis 2019 d'un projet de loi fédérale, l'Algorithmic Accountability Act, qui a vocation à lutter contre les biais des algorithmes via des obligations de transparence et un contrôle par la FTC, l'autorité de régulation du commerce américaine. À mon sens, il serait intéressant d'avoir une réglementation sur les algorithmes ou sur l'IA ayant un champ d'application global également en Europe. » L'Algorithmic Accountability Act pourrait d'ailleurs devenir plus contraignant pour les entreprises américaines, avec une nouvelle version du projet de loi déposée au Sénat et à la Chambre des représentants. L'objectif est notamment de veiller à ce que les évaluations d'impact mettent les consommateurs au premier plan. La Commission européenne s'engage, elle aussi, dans la voie législative avec une proposition de règlement présentée le 21 avril 2021 : l'Artificial Intelligence Act (AI Act).

« La feuille de route initiale était assez ambitieuse, puisque l'idée était de réguler l'ensemble des systèmes algorithmiques de manière large, rapporte Caroline Lequesne Roth. Dans le texte aujourd'hui en discussion, le champ d'application semble se réduire dans le cadre des négociations et certainement du lobbying qui est exercé. Une ONG en particulier, Algorithm Watch, s'est intéressée à l'application éventuelle de la législation auprès des administrations, notamment sur les systèmes de scoring qui étaient utilisés par les administrations sociales. Ce qui est surprenant, c'est qu'au terme de la seconde mouture – qui n'est pas officielle –, il semblerait que l'Union européenne écarte les systèmes qui impliquent une intervention humaine. Cela signifie que seuls les systèmes auto-apprenants entreraient dans le cadre de la définition. Typiquement, dans le cas du système autrichien sur le score d'employabilité conçu par l'agence pour l'emploi pour déterminer l'accès à la formation, nous estimons qu'il tomberait sous le coup de l'AI Act précisément parce que la participation humaine est quasi nulle. A contrario, le logiciel de la Chaf qui calcule les droits ne serait pas concerné, car le paramétrage a été fait par un humain. Ainsi, des systèmes très proches n'entreraient pas de la même

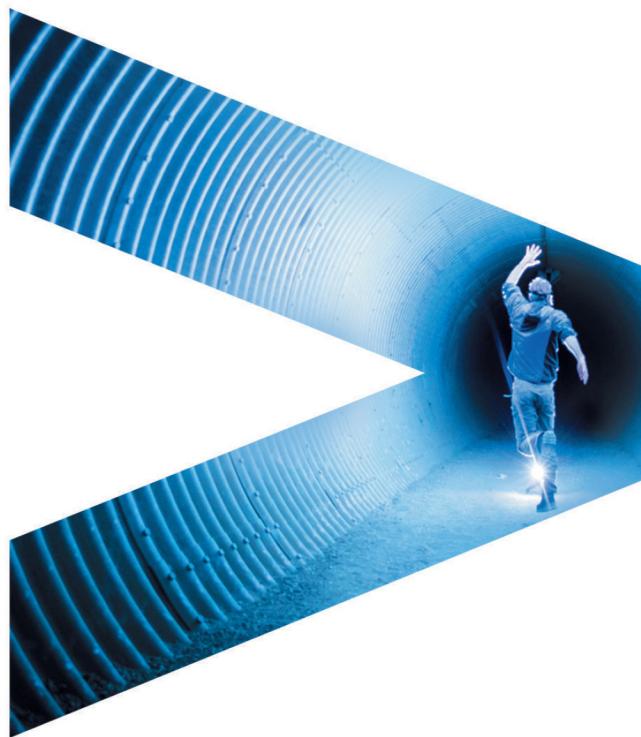
30

« Les algorithmes vont devenir si importants qu'il serait utile qu'un cadre réglementaire harmonisé aborde le sujet dans tous les domaines où ils interviennent. »



Benjamin ZNATY

Avocat du cabinet Taylor Wessing



Transformer l'actuariat avec les meilleurs experts

En alliant technologie et ingéniosité humaine, nous nous appuyons sur la puissance du changement pour créer une valeur à 360°. Nous donnons la priorité aux besoins de votre entreprise, en créant des solutions adaptées à votre secteur.

Place au changement

accenture

« La confiance doit se bâtir sur le terrain, de façon juridique et démocratique. »



Caroline LEQUESNE ROTH
Maître de conférences
en droit public

manière dans l'AI Act. Mais le texte n'est pas arrêté.» L'AI Act est fondé sur une approche par les risques, il distingue ainsi les IA à risque inacceptable, à haut risque, à risque limité et à risque minimal. « Celles à risque inacceptable seront interdites au sein de l'Union, c'est le cas de la notation sociale. C'est aussi le cas, avec d'importantes exceptions, de l'usage de la reconnaissance faciale en temps réel, décrypte Caroline Lequesne Roth. Une importante marge de manœuvre est laissée aux États sous couvert de protection des intérêts nationaux. Les IA à risque élevé concentrent l'effort régulatoire.»

« Les systèmes d'IA à haut risque liés au transport, aux dispositifs de santé, ceux qui sont déjà sous contrôle pour obtenir une certification et ainsi le marquage "CE" vont être très contrôlés, et ceci avant leur commercialisation ou mise en exploitation, confirme Philippe Besse. Pour d'autres systèmes jugés à haut risque (en justice, police, assurance...), il y a l'obligation de rédiger une documentation pour obtenir la certification, mais les contrôles ne se feront éventuellement que post commercialisation si les moyens suffisants sont alloués.» « Je crois qu'il est très important de renforcer les analyses d'impact dans le sens des droits fondamentaux, en faisant de cet outil un véritable levier de contrôle, ajoute Caroline Lequesne Roth. Il est important de réfléchir plus largement aux conditions d'usage de ces systèmes algorithmiques et de ce que signifie la garantie humaine. La confiance doit se bâtir sur le terrain, de façon juridique et démocratique. Il est indispensable d'avoir un État de droit fort, une prise en compte de l'humain et une participation citoyenne à la démocratisation de ces systèmes. Avec, notamment, la politique d'open data. Nous faisons face à une évolution des droits algorithmiques où l'on invite les citoyens à participer au contrôle démocratique C'est une voie qui mérite d'être approfondie. » L'avenir des algorithmes ne s'écrira pas sans l'humain.

Dossier réalisé par Coralie BAUMARD

32

actuariat

data science

gestion des risques

protection sociale

assurance

//galea

25 rue de Choiseul - 75002 Paris
01 43 22 11 11
contact@galea-associes.eu
www.galea-associes.eu



**CONSEIL POUR
LES ENTREPRISES
ET LES ORGANISMES
D'ASSURANCE**



#CONFORMITÉ CERTIFICAT MANAGEMENT DES RISQUES DE NON-CONFORMITÉ POUR ASSURANCE

Une formation conçue par les assureurs, pour les assureurs.



9 journées de formation

Accessible 100% en distanciel ou en présentiel

Début de la formation : 11 avril 2022



Christelle PERRI

***"Cette formation m'a beaucoup apporté,
j'espère qu'elle sera un bon tremplin."***



Yannick QUERE

***"Je vous félicite pour la qualité du contenu,
des intervenants mais également des stagiaires.
(...) Ces journées de formation ont été pour moi
de vrais moments d'enrichissement professionnel.
Bonne continuation à ce certificat
que je ne manquerai pas de promouvoir."***



Matthieu Gray

***"Cette formation était vraiment très bien organisée (...)
en plus d'être de très grande qualité. Je considère
que j'ai fait un très bon choix en décidant d'y participer.
Elle m'a beaucoup apporté [...], nous avons trouvé dans
mon groupe beaucoup d'applications pratiques."***



institut du risk
management

