# Réseaux de neurones et responsabilité : quand l'information se trouve dans des couches cachées

8 octobre 2019

## Auteur

Eric Lavallée

Associé, Agent de marques de commerce Associé, et Avocat

**Bon nombre des techniques d'apprentissage automatique les plus avancées ont recours à des réseaux de neurones artificiels, lesquels permettent aux systèmes « d'apprendre » des tâches en assimilant des exemples, et ce, sans avoir été spécialement programmés pour exécuter de telles tâches.**

Bien que les réseaux neuronaux ne datent pas d'hier, l'avènement de l'apprentissage profond[1] et de la capacité des ordinateurs à traiter rapidement de grandes quantités de données a mené à la mise au point d'un éventail de solutions d'apprentissage automatique touchant divers aspects de la vie. De la reconnaissance d'images au traitement de données financières, l'apprentissage automatique devient de plus en plus omniprésent.

Selon une perspective mathématique, les réseaux neuronaux modernes utilisent presque toujours des « couches cachées ». Celles-ci traitent l'information de la couche d'entrée vers la couche de sortie du réseau neuronal. Aucune tâche et aucun poids n'est expressément attribué par un programmeur aux neurones des couches cachées. Or, en règle générale, il n'existe aucune façon directe de savoir comment l'information est traitée dans ces couches.

En termes simples, la plupart des techniques d'apprentissage automatique actuelles ont recours à des méthodes dont le fonctionnement ne permet pas aux opérateurs humains de connaître l'intégralité des étapes du processus. Par conséquent, les systèmes qui usent de ces méthodes poseront de nouveaux défis juridiques aux avocats.

Des chercheurs étudient cette question depuis plus d'une décennie[2], mais ils ne sont toujours pas parvenus à donner des réponses définitives.

Les questions de cette nature sont au cœur des débats juridiques actuels. Dans une décision de la Cour suprême des États-Unis portant sur une affaire très médiatisée de découpage électoral

partisan[3], des préoccupations quant à l'apprentissage automatique ont été soulevées par la juge dissidente. Cela n'a rien d'étonnant, compte tenu du fait que les parties ont présenté aux tribunaux inférieurs la preuve portant sur les *méthodes de Monte-Carlo par chaînes de Markov*[4], lesquelles, tout comme les réseaux neuronaux, ne permettent pas à l'opérateur humain de savoir en détail comment chaque donnée entrée affecte les résultats.

Dans certains pays, comme aux États-Unis, un utilisateur d'une technologie peut être en mesure de rejeter des demandes de divulgation d'algorithmes et de détails concernant le processus d'apprentissage automatique de ladite technologie en soutenant que ces renseignements sont protégés en tant que secrets commerciaux du fournisseur de la technologie visée[5]. Malgré tout, il pourrait être nécessaire de communiquer certaines informations, comme les résultats du processus d'apprentissage automatique dans différentes situations, pour démontrer sa fiabilité et son caractère adéquat.

Un tel argument pourrait ne pas être valable dans d'autres pays. Par exemple, en France, le Conseil constitutionnel a récemment autorisé l'administration publique à avoir recours à un processus algorithmique dans le cadre de la prise de décision, mais *seulement* si elle est en mesure de communiquer, en détail et sous une forme intelligible, la façon dont ce processus algorithmique prend ses décisions[6]. D'un point de vue informatique, il est difficile de satisfaire à ces exigences lorsque l'on considère le concept des couches cachées.

Plus important encore, une personne pourrait vouloir communiquer la façon dont une technologie d'apprentissage automatique l'a aidée à prendre une décision afin de prouver qu'elle a agi comme il se doit. Par exemple, pour éviter que leur responsabilité professionnelle soit engagée, certains professionnels de la santé peuvent être appelés à expliquer comment l'apprentissage automatique les a guidés dans leur prise de décision.

Une décision récente de la Cour du banc de la Reine de l'Alberta[7] relative à la responsabilité professionnelle des médecins démontre à quel point une telle preuve peut être complexe. Dans celle-ci, l'un des facteurs permettant de déterminer la responsabilité du médecin était le poids fœtal et les différentes formules qui auraient pu être utilisées pour l'établir.

La Cour a déclaré ce qui suit : « […] l'expertise requise porterait sur le développement des algorithmes utilisés pour les calculs automatisés de l'indicateur composite du poids à la naissance en considérant les recherches empiriques concernant les poids réels à la naissance et les variables ou les facteurs utilisés pour calculer l'indicateur composite du poids à la naissance. Aucune personne ni aucun groupe de personnes ayant une telle expertise n'a témoigné. Je ne tire pas de conclusion en ce qui concerne les calculs du rapport d'échographie du mois de février se basant sur des formules et des estimations du poids différentes. » (*Traduction non officielle*).

Pour les développeurs et les utilisateurs de technologies d'apprentissage automatique, il est donc important de consigner les informations utilisées pour former leur algorithme, la façon dont le système a été mis en place et le raisonnement derrière le choix des diverses méthodes technologiques utilisées pour l'apprentissage automatique.

Les informaticiens qui ont développé des applications visant des domaines précis devraient travailler en étroite collaboration avec des experts dans ces domaines afin de veiller à ce que les données utilisées pour former l'algorithme soient adéquates et que l'algorithme en résultant soit fiable.

Dans certains cas, il pourrait même être nécessaire de développer des technologies supplémentaires servant à assurer le suivi de l'information qui parcourt le réseau neuronal et à sonder les couches cachées[8].

**Que retenir?**

Les risques liés à l'usage d'un système incorporant de l'apprentissage automatique doivent être évalués dès sa conception. Il est recommandé de consulter un avocat dès ce moment pour bien orienter le projet.
Lorsque c'est possible, il faut orienter les choix technologiques vers des approches robustes dont les résultats seront aussi stables que possible.
Il est important de documenter ces choix technologiques et l'information utilisée lors du développement d'algorithmes d'apprentissage automatique.
Les contrats entre les concepteurs et les usagers des technologies doivent clairement attribuer les risques entre les parties.

---

1. Voir, en particulier : Rina Dechter (1986). *Learning while searching in constraint-satisfaction problems*. Université de Californie, Département de sciences informatiques, Cognitive Systems Laboratory, 1986.; LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey (2015). « Deep learning ». *Nature*. 521 (7553): 436–444.
2. Par exemple : Matthias, Andreas. « The responsibility gap: Ascribing responsibility for the actions of learning automata. » *Ethics and information technology* 6.3 (2004): 175-183; Singh, Jatinder, et al. « Responsibility & machine learning: Part of a process. » Accessible au SSRN 2860048 (2016); Molnar, Petra et Lex Gill. « Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System. » (2018).
3. *Rucho v. Common Cause*, No. 18-422, 588 U.S. ___ (2019).
4. 279 F.Supp.3d 587 (2018).
5. *Houston Fed. of teachers v. Houston Independent*, 251 F.Supp.3d 1168 (2017); *Brennan Ctr. for Justice at New York Univ. Sch. of law v. New York City Police Dept*. 2017 NY Slip Op 32716(U) (NY Supreme Court).
6. Décision no 2018-765 DC datée du 12 juin 2018 (*Loi relative à la protection des données personnelles*).
7. *DD v. Wong Estate*, 2019 ABQB 171.
8. Par exemple : Graves, Alex, Greg Wayne et Ivo Danihelka. Neural Turing Machines. arXiv:1410.5401, [cs.NE], 2014.