

# Les bugs de l'intelligence artificielle

Par David Larousserie

Publié le 25 février 2019 à 19h02 - Mis à jour le 26 février 2019 à 16h03

## [Les bugs de l'intelligence artificielle](#)

- [Réservé à nos abonnés](#)Partage
- Partage désactivé
- Partage désactivé
- Partage désactivé

**ENQUÊTE** Les succès de l'IA dans les domaines les plus divers, de la conduite automatique au diagnostic médical,

**feraient presque oublier qu'elle reste faillible.**

Côté face, tout va bien pour l'intelligence artificielle, surtout pour son volet dit d'apprentissage automatique (ou statistique). Depuis le début des années 2010, elle vole de succès en succès. Ses algorithmes battent les humains au jeu de go, aux échecs, au poker, au casse-briques ou aux jeux d'arcade. Ils conduisent des voitures autonomes sur des milliers de kilomètres (presque) sans accident mortel. Commandés par la voix, ils lancent des flux radio, des achats en ligne, des bulletins météo... Ils diagnostiquent mieux que les médecins.

Côté pile, c'est moins riant. « *Je ne ferais pas confiance à une très grande partie des découvertes en cours qui recourent à de l'apprentissage automatique appliqué à de*

*grands ensembles de données* », a prévenu le 15 février Genevera Allen, de l'université Rice, lors du congrès annuel de l'Association américaine pour l'avancement de la science, comme le rapporte le *Financial Times*. Dans leurs articles de recherche ou dans des tribunes publiées dans les médias, des spécialistes parlent d'« *intelligence artificielle trompée* », de « *failles dans l'apprentissage automatique* », de « *crise* », de « *trouble profond* »...

Beb-Deum

C'est l'intelligence artificielle même qui est touchée au cœur : des systèmes de reconnaissance d'images prennent des vessies pour des lanternes, des « pirates » attaquent les algorithmes et les font dérailler, les corrections des biais touchant ces algorithmes mettent à l'épreuve les statisticiens et les plus

visionnaires réfléchissent à la manière de rendre plus intelligentes ces machines qui, il faut bien le dire, sont encore assez bêtes.

## **Apprendre par l'exemple**

Evidemment, corriger ces défauts devient prioritaire, d'autant que l'avenir promet des applications bien plus critiques que de jouer au go ou distinguer un chaton d'un lapereau : diagnostic médical, conduite de véhicules, aide à la décision pour des verdicts judiciaires, pour l'attribution de crédits... Alors dans les labos, en même temps que certains bricolent leur système pour être meilleurs que les concurrents, d'autres ne parlent que de quête de « robustesse », « fiabilité », « certification », « reproductibilité », « confiance », « équité »...

Avant de passer en revue ces défauts et les corrections que tente d'y apporter la

communauté scientifique, rappelons brièvement ce que sont ces systèmes d'apprentissage automatique.

Comme leur nom l'indique, il s'agit, tel un enfant, d'apprendre par l'exemple. On montre au programme informatique des cas particuliers afin qu'il ajuste ses paramètres pour coller à la bonne réponse fournie (par exemple la nature de l'objet dans l'image). Puis, magie de la technique, même pour une image inconnue, le système généralisera et donnera la bonne réponse. « Petit » détail, la fonction mathématique qui permet de passer des images aux réponses contient des millions, voire des milliards de paramètres. Historiquement, ces paramètres ont été comparés à des neurones, d'où le nom de « réseaux de neurones », dont une variante est appelée *deep learning*. Une autre façon de voir ces algorithmes est de se les

imaginer capables, dans un espace immense où chaque point serait une image, de tracer une frontière géométrique séparant les objets de nature différente.

**Il vous reste 80.31% de cet article à lire.  
La suite est réservée aux abonnés.**