Explicabilité : le mauvais procès fait à l’intelligence artificielle

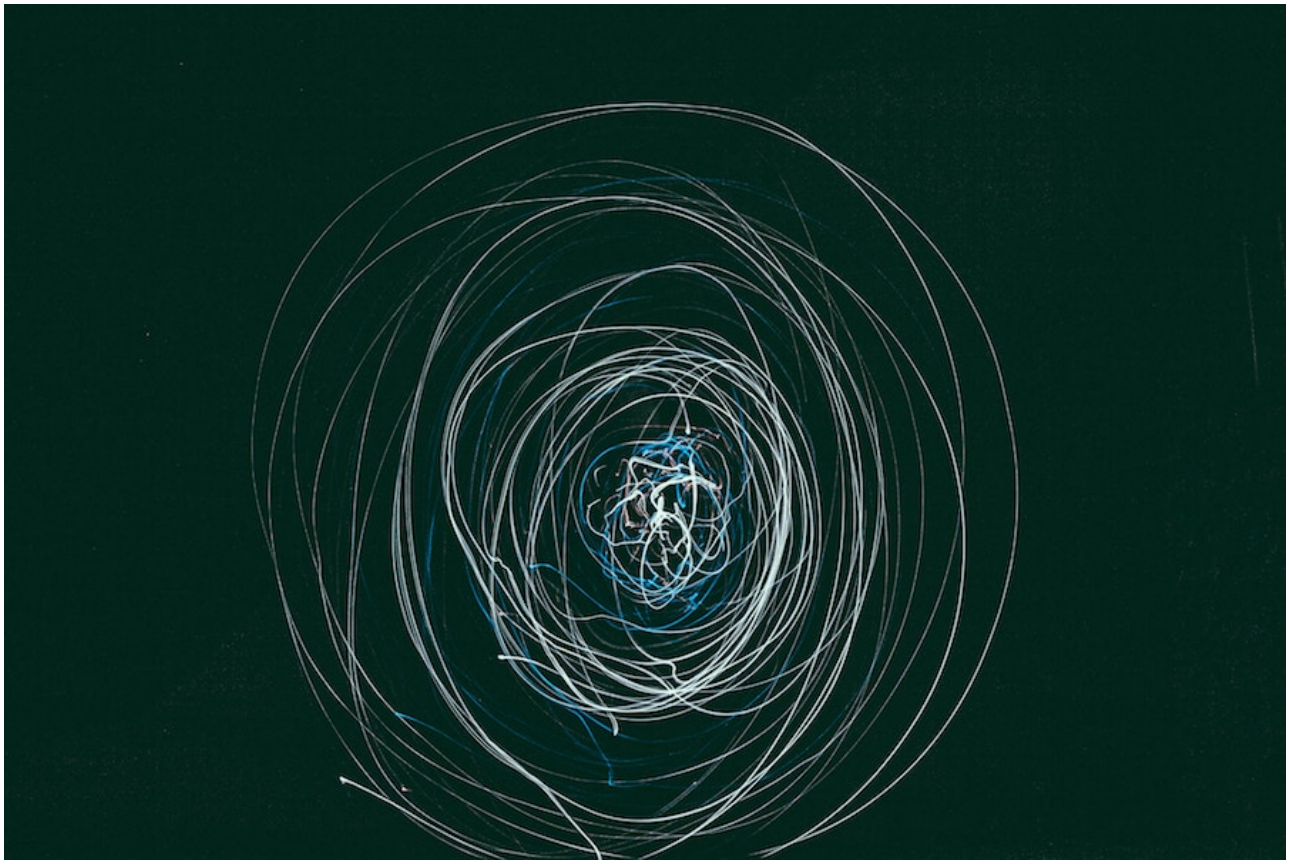
Comprenons-nous vraiment les rouages de l'intelligence artificielle ? Maitrisons-nous vraiment ce qu'il se passe avec cette technologie ? Manuel Davy, Président et fondateur de Vekia, fait le point.

Temps de lecture : minute

10 juillet 2018

"*Personne ne sait vraiment comment les algorithmes les plus avancés font ce qu'ils font. Ça pourrait être un problème*." Cet extrait tiré de "*The Dark Secret at the Heart of AI*", la revue technologique du MIT, rend compte d’une des entraves à l’acceptation de l’IA : son explicabilité. C’est-à-dire sa capacité à expliquer par quel raisonnement certains algorithmes d’apprentissage parviennent à de bons résultats. Preuve que ce sujet vieux comme l’IA reste éminemment d’actualité et dépasse les frontières du monde scientifique, le rapport Villani allait jusqu’à qualifier l’explicabilité d’ "*enjeu démocratique*".

Le mutisme des algorithmes tient à la complexité du lien entre les données analysées par l’intelligence artificielle et la décision prise. Une complexité qui, d’ailleurs, échappe à la machine elle-même. Dans son ouvrage "*Des robots et des hommes*", paru l’an dernier, Laurence Devillers, chercheuse au CNRS, rappelait qu’ "*il faut garder à l’esprit que ces technologies apprennent sans comprendre donc qu’elles décident sans comprendre*". Les réseaux de neurones du *deep learning* en sont un parfait exemple. Constitués de couches qui transforment successivement les données d’entrée jusqu’à calculer finalement une décision, les données transformées par les couches les plus profondes n’ont aucun sens pour l’humain. La décision prise est donc inexplicable.



À lire aussi
L’Intelligence Artificielle, un concept (encore) à définir

Si ce manque de transparence embarrasse les scientifiques, il constitue surtout une source d’inquiétude pour les opinions publiques. A l’ère de la transparence, de l’horizontalité et du rendre compte, cette opacité bouscule notre fonctionnement. Toutefois refuser l’intelligence artificielle reviendrait à mettre à mal notre cohérence.

"L’intelligence artificielle n’est pas la seule réalité que nous ne parvenons pas à appréhender"

Il est vrai qu’en quelques années à peine, l’intelligence artificielle a fait irruption dans notre quotidien, suscitant beaucoup de questions. Pourtant, il ne s’agit pas de la seule réalité que nous ayons du mal à appréhender.

Combien d’entre nous connaissent les mécanismes physiques, chimiques et biologiques à l’œuvre dans le langage par exemple ? Très peu. Et pourtant cela ne nous empêche pas de parler. Nous ne savons pas toujours expliquer les dysfonctionnements de ce mécanisme hautement complexe, et pourtant nous vivons bien avec.

À lire aussi
Comment l’intelligence artificielle va rendre notre vie terriblement ennuyeuse

A l’évidence, l’humain est capable d’accepter ce qu’il ne comprend pas. Toutefois, comme pour toutes les innovations majeures, l’intelligence artificielle ne bénéficie pas de la même tolérance pour le moment. Sans doute parce qu’elle se propose d’agir – sans s’expliquer – sur des aspects du quotidien en lien avec notre santé et notre sécurité. Mais là encore l’analogie est de mise.

Lorsqu’un médecin pose un diagnostic, nous ne connaissons ni ne

comprenons la totalité des informations l'ayant conduit à sa décision ; et quand bien même nous avons accès à des résultats d'analyse, nous ignorons le fonctionnement de la méthode de biologie moléculaire utilisée. De même lorsque nous roulons en voiture nous sommes nombreux à ignorer les mécanismes profonds du système de freinage impliquant le liquide de frein, le frein à tambour, les plaquettes ou le disque. Pourtant cela ne nous conduit pas (c'est le cas de le dire) à douter de notre voiture.

A l'évidence c'est la confiance liée à l'habitude ou à l'autorité morale qui nous incite à accepter ce que nous ne comprenons pas entièrement. Y compris dans des domaines aussi sensibles que ceux que nous venons de citer qui engagent notre santé ou notre sécurité.

Des initiatives qui rassurent

En se déployant avec succès sur des applications nouvelles, l'intelligence artificielle multiplie les gages de confiance. Par exemple, en mai dernier, des chercheurs allemands ont conçu un algorithme pour dépister le cancer de la peau. Ses performances ont été comparées à celles de 58 médecins spécialistes originaires de 17 pays différents. Ces derniers ont identifié 87% des mélanomes qui leur étaient présentés. Le robot a fait mieux avec 95% de mélanomes détectés.

À lire aussi
IA : la guerre des talents

Toutefois, ces réalisations ne suffisant pas encore à rassurer complètement, les chercheurs doivent faire encore progresser l'IA. Plusieurs pistes sont envisageables : recourir à des formules mathématiques simplifiées expliquant " dans les grandes lignes " les liens entre les variables d'entrée et de sortie, former des équipes de data scientists spécialisées dans l'analyse de cas pour fournir des explications ou mettre au point des chatbots qui répondent sans se lasser à des questions sur un résultat et qui permettent à l'utilisateur d'explorer à l'infini les détails d'un calcul.

Ces initiatives seront, sans doute possible, couronnées de succès. C'est aussi pour cela que l'intelligence artificielle fera toujours plus présente dans notre quotidien.

Article écrit par Manuel Davy