

➤ dépourvu de toute connaissance en géométrie et en physique. Le programme a réussi à «découvrir» plusieurs lois (les Hamiltoniens, les Lagrangiens, la conservation du moment...). En d'autres termes, l'algorithme s'est mué en un physicien ! Dans ce cas, les lois identifiées étaient connues, mais pourquoi ne pas imaginer qu'un algorithme puisse identifier de nouvelles lois ?

Le *big data* peut également aider à la compréhension d'événements impliquant de très grands nombres de variables quantifiables, tels les phénomènes météorologiques ou climatiques, les comportements électoraux, l'usage des réseaux sociaux...

FUYEZ L'HÔPITAL !

Toutefois, ce transfert vers les *big data* n'est pas sans risque. De fait, il est aussi possible que nous nous égarions dans l'identification de multiples corrélations, pas forcément bien interprétées. Or une corrélation n'est pas une relation de cause à effet. Elle marque simplement le fait que deux grandeurs semblent dépendre l'une de l'autre, au sens où si l'une augmente, l'autre augmente ou décroît «de la même façon».

Par exemple, il a été récemment démontré dans une université américaine que les étudiants qui sont les plus gros consommateurs d'alcool ont tendance à avoir des résultats scolaires moins bons que les autres. L'alcool est-il la cause directe de cette baisse de niveau ? N'allons pas trop vite en besogne. On peut certes imaginer que l'alcool rende stupide, mais d'autres hypothèses sont aussi envisageables : les étudiants qui boivent de l'alcool seraient à la base de moins bons étudiants ; ou bien ils seraient moins attentifs que les autres les lendemains de soirées arrosées, expliquant leurs déboires aux examens ; ou bien encore, ils boiraient plus que de raison afin de calmer leur peur d'échouer ou pour noyer des chagrins n'ayant rien à voir avec leurs études...

Ainsi, une corrélation manifeste n'est pas nécessairement la manifestation d'une relation de causalité : ce n'est pas parce qu'il y a des grenouilles après la pluie qu'on a le droit de dire qu'il a plu des grenouilles. Mais il arrive très souvent que nous confondions les deux notions, à la manière d'un Coluche conseillant de ne jamais aller à l'hôpital au motif qu'on y meurt en moyenne plus souvent que chez soi... Plusieurs sites recensent ce type de corrélations (voir la figure page ci-contre) relevant de l'effet cigogne : le taux de natalité augmentant comme le nombre de nids de cigognes, on en déduit que les cigognes apportent les bébés ! Il y a matière à sourire, mais le problème n'en demeure pas moins sérieux.

Il y a bien la possibilité qu'avec le *big data*, au lieu de théoriser, nous cédions aux facilités

de l'induction et délaissions le «geste théorique» de type galiléen, celui qui consiste à énoncer des hypothèses portant bien au-delà des données disponibles.

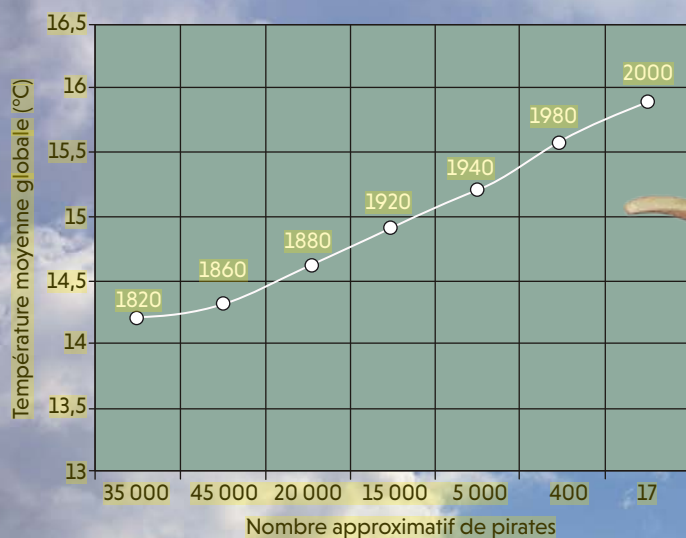
À cet égard, le cas d'Albert Einstein est exemplaire. En 1915, il publiait la théorie de la relativité générale, une conception révolutionnaire de la gravitation, alors qu'il n'avait que très peu de données sur l'Univers à sa disposition : on ignorait, par exemple, pourquoi les étoiles brillent, que d'autres galaxies hormis la

N'allez jamais à l'hôpital, on y meurt en moyenne plus souvent que chez soi

notre existent, que l'Univers est en expansion, etc. Cela n'a pas empêché les équations d'Einstein de parfaitement s'accommoder de la quantité gigantesque de données recueillies depuis un siècle par les télescopes et les satellites. En outre, elles constituent les fondations d'une véritable cosmologie scientifique, capable d'envisager l'Univers en tant qu'objet physique, doté de propriétés qui le caractérisent «lui-même».

Elles ont même permis de prédire l'existence des ondes gravitationnelles un siècle avant leur première détection, en septembre 2016. Une théorie peut donc non seulement enrichir l'univers des données, mais également agir comme un «treuil ontologique» capable de faire apparaître de nouveaux éléments de réalité. En d'autres termes, la théorie en «dit plus» que les données, notamment par le fait qu'elle explicite des lois que les données n'illustrent jamais que de façon partielle.

Imaginons maintenant que les choses se soient passées dans l'ordre inverse, c'est-à-dire que nous ayons commencé avec toutes les données dont nous disposons aujourd'hui, mais sans avoir à notre disposition la théorie de la relativité générale. Pourrions-nous, par une sorte d'induction théorique permettant de passer des données aux lois, découvrir les



équations d'Einstein? Rien n'est moins sûr, même si, nous l'avons vu, certaines lois physiques simples ont pu être récemment «redécouvertes» à partir de l'analyse d'un ensemble de données expérimentales.

La réponse d'Einstein à cette question aurait en tout cas été négative, du moins si l'on en croit la lettre qu'il écrivit un jour à son grand ami Maurice Solovine: «Aucune méthode inductive ne peut conduire aux concepts fondamentaux de la physique. L'incapacité à le comprendre est la plus grave erreur philosophique de nombreux penseurs du XIX^e siècle.»

Quelques années plus tard, lors d'une conférence qu'il donna à Oxford avant de quitter définitivement l'Europe pour l'Amérique, il précisera quelle était selon lui la base de l'invention des idées scientifiques: «L'expérience peut bien entendu nous guider dans notre choix des concepts mathématiques à utiliser, mais il n'est pas possible qu'elle soit la source d'où ils découlent. C'est dans les mathématiques que réside le principe vraiment créateur. En un certain sens, donc, je tiens pour vrai que la pensée pure est compétente pour comprendre le réel, ainsi que les Anciens l'avaient rêvé.»

En la matière, les prochaines décennies, toutes gorgées d'intelligence dite «artificielle» et de «*very big data*», viendront-elles contredire la pertinence de cet avis? Il est raisonnable d'en douter, pour au moins deux raisons.

La première tient à ce que l'intelligence artificielle porte bien son nom: elle n'est qu'une intelligence fabriquée au moyen de techniques informatiques, de sorte que rien

Selon le pastafarisme, une parodie de religion parfois considérée comme une version moderne de la théière de Russell (une métaphore qui illustre le fait que c'est au croyant de prouver les bases «*invérifiables*» de la religion), le déclin de la piraterie est la cause du réchauffement climatique, car la température moyenne sur la Terre est corrélée, de façon inverse, à la population de pirates.

ne permet de penser qu'elle soit elle-même «pensante». Pour doter les machines d'une pensée autonome, analogue à la pensée humaine, il faudrait déjà comprendre en détail les mécanismes de cette dernière, ce qui est loin d'être acquis, et aussi en donner une définition qui soit à la fois opératoire et exhaustive. Classer des milliers de photographies mieux et plus vite que n'importe quel humain est une chose, mais inventer des concepts et les associer en est une tout autre...

APPELONS UN CHAT UN CHAT, MAIS COMMENT ?

La seconde tient en ce que les techniques de l'intelligence artificielle, tel le *deep learning* (voir *La révolution de l'apprentissage profond*, par Y. Bengio, page 42), fonctionnent comme des boîtes noires. Lorsque je vois un chat et reconnais aussitôt qu'il s'agit bien d'un chat, j'effectue une opération qui n'est pas de même nature que celle d'un logiciel d'intelligence artificielle: lui constate que l'objet qui lui est présenté ressemble à d'innombrables autres chats dont on lui a préalablement fourni les images pour «l'entraîner», mais sans que quiconque puisse déterminer comment il effectue ce constat, ni indiquer quelles sont les ressemblances qui font «tilt» pour lui.

Il est donc difficile d'imaginer que l'intelligence artificielle puisse élaborer des théories physiques, et encore moins, si d'aventure elle le pouvait, que celles-ci puissent nous être intelligibles. Jusqu'à preuve du contraire, physicien demeure un métier d'avenir. ■

BIBLIOGRAPHIE

G. BERRY,
L'hyperpuissance de l'informatique : Algorithmes, données, machines, réseaux, Odile Jacob, 2017

M. SCHMIDT ET H. LIPSON,
Distilling Free-Form Natural Laws from Experimental Data, *Science*, vol. 324, pp. 81-85, 2009.