

Dans cet article, le physicien Hubert Krivine – auteur de nombreux ouvrages, dont [L'IA peut-elle penser ?](#) – analyse la manière dont fonctionne [GPT-3](#) (Generative Pre-trained Transformer 3) et, à partir de réflexions sur les sciences modernes qu'il poursuit depuis plusieurs livres, revient sur les limites de l'intelligence artificielle.

Dans une Lettre à Kepler du 19 août 1610, Galilée se moque de ceux qui pensent :

[...] chercher la vérité ni dans le monde, ni dans la nature, mais (je cite leurs propres paroles) dans la confrontation des textes.

Galilée visait les lecteurs qui croyaient trouver dans le Livre Saint la réponse à toutes les questions. Pour Galilée et les savants modernes, seule la théorie et l'expérience permettent d'approcher ce qu'il est convenu d'appeler une vérité scientifique qui n'est évidemment jamais absolue et fixée^[1]. Mais il ne suffit pas d'observer le monde en accumulant des données. Il faut aussi savoir que^[2] :

La philosophie est écrite dans ce livre gigantesque qui est continuellement ouvert à nos yeux (je parle de l'Univers), mais on ne peut le comprendre si d'abord on n'apprend pas à comprendre la langue et à connaître les caractères dans lesquels il est écrit. Il est écrit en langage mathématique.

En d'autres termes, même si on peut questionner l'aspect un peu réducteur du « langage mathématique », il faut raisonner, nous dit Galilée.

Quelles que soient ses spécificités, l'IA moderne (à la différence des systèmes experts) travaille par induction. C'est à dire qu'elle se nourrit de données (les *data*) pour en singulariser des corrélations et en tirer des prévisions. C'est ainsi que travaillaient les savants jusqu'à la Renaissance. Avec des résultats remarquables en astronomie, agriculture, architecture, mais quelque fois tout simplement faux ; par exemple l'affirmation que les corps doivent tomber d'autant plus vite qu'ils sont lourds, que la Terre est fixe ou que la génération spontanée existe.

Est venue ensuite l'idée que, là où c'était possible, la théorie – souvent appuyée par les mathématiques – pouvait être une arme bien plus féconde. Mais, depuis le début du XX^{ème} siècle, on a compris que, dans certains cas, même la connaissance de lois ne permet pas de prévoir (c'est l'effet papillon plus savamment appelé chaos déterministe)^[3] ou alors qu'elles sont si complexes que les approximations nécessaires pour les appliquer peuvent les rendre fautives. L'idée est alors que le retour à la méthode inductive appuyée cette fois-ci par les *milliards* de données fournies par les *Big data* pourrait être réhabilitée.

Grosso modo, le raisonnement inductif consiste à considérer que si des événements se sont produits n fois, la probabilité qu'ils se reproduisent $n+1$ fois est d'autant plus élevée que n est grand. C'est le « jamais deux sans trois ! ». On sait qu'il est trompeur : vos parents et vous-même ont vécu jusque- là et ça va continuer... Mais il est très employé dans la vie courante.

Le physicien (et homme d'affaire) Chris Anderson publie dans le revue *wired*, en juillet

2008, un article célèbre au titre explicite « La fin de la théorie : le déluge de données rend la méthode scientifique obsolète ». On y trouve qu’

avec suffisamment de données, les nombres parlent d’eux-mêmes,

et plus loin que :

La corrélation supplante la causalité, et la science peut avancer même sans modèle cohérent, sans théorie unifiée voire sans aucune explication mécaniste du tout.

Pour lui, «tous les modèles sont faux» et souvent pollués d’idées préconçues alors que les bases de données, à condition d’être suffisamment gigantesques ne sauraient mentir. Ce que ne comprend pas Anderson, c’est que :

1- Ce qu’il appelle « données » sont en vérité des « prises ». C’est-à-dire résultent d’un travail conscient ou – le plus souvent – inconscient de sélection. Et alors la « loi des grands nombres » appliquée à des données biaisées n’immunise plus. Et au contraire propage l’illusion de l’objectivité. Les données « brutes » n’existent pas.

2- De plus, les théories scientifiques les plus fécondes comme par exemple la mécanique quantique ou la relativité ne découlent absolument pas de l’exploitation de big data. Elles en seront au contraire une source. Par exemple les ondes gravitationnelles ne sont devenues des « données » observables que parce que la théorie d’Einstein les avait prévues. On ne les aurait jamais trouvées sinon. Les exemples plus anciens sont multiples. Arago a joliment écrit que :

M. Le Verrier a aperçu le nouvel astre sans avoir besoin de jeter un seul regard vers le ciel ; il l’a vu au bout de sa plume^[4].

3- Une théorie est infiniment plus riche que la synthèse de nos connaissances ; c’est elle qui permet la prévision de choses inouïes au sens étymologique (Comme les ondes électromagnétiques, le boson de Higgs, la matière noire...).

4- Cette capacité du cerveau à produire des théories reste certainement toujours mystérieuse. Elle est en tout cas certainement reliée au fait que le cerveau est une machine très particulière ; elle est « bio-logique », c’est-à-dire en rapport étroit avec le corps qui le nourrit et en interaction permanente avec le reste de la société. Aucune IA – en tout cas aujourd’hui – n’a ces caractéristiques.

Ce très bref rappel des limitations de l’IA^[5] fait, venons-en à son utilisation avec GPT-3. GTP-3 est un traitement automatique de langue (TAL). Même s’il donne l’impression de répondre à des questions, il n’est pas du tout fait pour ça. Sur une base *gigantesque* de données, il cherche dès qu’un mot (ou une phrase) est proposée, l’occurrence la plus vraisemblable du mot suivant, puis il itère. Un peu comme votre téléphone portable le fait, mais ici sur une base de données des milliards de fois plus grande. Il a beaucoup de problèmes, mais la recherche du vrai est le cadet de ses soucis. Un exemple que je connais

assez bien : qui est l'auteur de ce papier ? La réponse a été :

Hubert Krivine est un médecin, ancien militant politique et écrivain français. Il est surtout connu pour son engagement politique dans les années 60 et 70 en France, comme membre de l'Organisation communiste internationaliste (OCI) et de Lutte Ouvrière. Il a été également membre du Parti de Gauche pendant quelques années. Il est aussi connu pour son engagement en faveur de la cause palestinienne et pour ses positions controversées en matière de politique étrangère. Il a écrit plusieurs livres sur les sujets politiques et historiques. Il est aujourd'hui retraité de la politique active mais continue de s'exprimer sur des sujets qui lui tiennent à cœur.

Pour quelqu'un ignorant tout de l'affaire, la réponse, écrite en bon français, est bluffante de véracité. Mais elle est simplement fantaisiste, elle met bout à bout des mots qui peuvent « aller ensemble » qui ont été pêchés par proximité dans des textes différents. C'est une source inépuisable et automatique de fake news. Sur la même requête faite un peu plus tard (les réponses ne sont jamais les mêmes) j'apprends que je suis mort en 2020. Ce que j'ignorais...

Wikipédia fonctionne, pourrait-on dire, à rebours de cette logique. Il est le fruit de connaissances (et pas de mots) collectées, en principe vérifiées et surtout il fournit ses sources. C'était un pari audacieux dans les années 2000 que de faire fonctionner une encyclopédie sans rédacteur attitré, mais reposant sur une collaboration autodisciplinée de ses lecteurs. Pari gagné : en nombre d'erreurs il le dispute aux grandes encyclopédies, comme l'Encyclopédia Britannica ou Universalis. Et surtout il se réactualise en permanence, ce que ne peut pas faire GPT-3. On nous promet une nouvelle version GPT-4, avec correction humaine des plus gros biais. Reviendrait-on alors à une logique Wikipédia ?... On verra.

Mais quel rapport avec ce vieux Galilée cité en prologue de ce papier ? C'est qu'au lieu de chercher la vérité « dans le monde et la nature », GPT-3 ne peut que se limiter à la chercher dans « les textes » déjà produits et même pire, dans les assemblages de mots fabriqués sur leur probabilité d'occurrence piquée un peu partout. La démarche scientifique interroge certainement la littérature. Pas pour y trouver une réponse, mais pour aller au-delà. Pour questionner intelligemment la nature, en s'aidant des connaissances acquises.

Bien entendu, nous avons tordu le bâton dans un sens, celui du bluff permis par GPT-3. Bluff qui nous a également estomaqué. Lire une copie d'histoire - relativement - bien rédigée mais totalement bidon d'un élève de Terminale, ou ne pas faire la différence entre un vrai discours de Ministre ou un fabriqué par la machine ne peut pas laisser de marbre.

Je suis évidemment convaincu de l'utilité décisive de l'IA dans bien des domaines, en tout cas dans les domaines qui peuvent être analysés « tout chose égale par ailleurs » (voir la note 5 en bas de page). Il est inutile de les énumérer dans ce court papier : la littérature en est saturée. Mais ses dangers souvent dénoncés se voient multipliés par l'usage des traitements automatique du langage qui en est un sous-produit spectaculaire certes, mais certainement pas le plus décisif. Peut-être pour le business, mais pas pour la pensée humaine.

Notes

^[1] Mais nous n'en connaissons, hélas, pas d'autres !

^[2] Dans une célèbre passage du *Saggiatore* (1623).

^[3] Nous laisserons de côté la mécanique quantique où les fameuses relations de Heisenberg sont souvent invoquées à tort et à travers.

^[4] La présence d'une planète inconnue pouvait expliquer les (petites) différences de trajectoire d'Uranus avec les prévisions de la théorie de Newton. Ce sera en 1846, la découverte de Neptune.

^[5] Le lecteur intéressé pourra lire les articles sur l'IA dans les n° de Contretemps.