

**Intelligence artificielle**  
**Jacques Malherbe**  
**Avocat (Simont Braun)**  
**Professeur émérite de l'UCLouvain**

### **A la recherche d'une définition**

L'intelligence artificielle fait irruption dans le monde informatique sous la forme d'un robot conversationnel ou chat box, développé par la société Open AI liée à Microsoft. Le magazine Der Spiegel la qualifie de nouvelle puissance mondiale. De bonnes émissions de la radio-télévision belge d'expression française nuancent quelque peu ce propos. Le concept est né dès les années 1950, avant même l'ordinateur et, en tous cas, l'internet. L'intelligence artificielle permet à un robot qui a absorbé des quantités énormes de données, par exemple la totalité de l'encyclopédie de l'internet Wikipédia, laquelle ne représente toutefois qu'un pourcent des textes absorbés, de restituer des textes en réponse à des demandes qui lui sont faites. Il s'agit toutefois non d'un module de connaissances mais d'un module de vraisemblance. Le système a procédé par apprentissage et peut donc restituer, en déterminant le sens d'un mot d'après son contexte, la réponse la plus vraisemblable à une question ou l'exposé d'une théorie ou encore le rappel de faits historiques d'après ce qu'il a appris. Il s'agit d'un système de neurones artificiels qui possède 175 milliards de paramètres qu'il peut moduler.

Toutefois, il ne comprend pas ce qu'il dit mais fonctionne par simple inférence statistique. Il ne pourrait donc, contrairement à ce qu'on croit souvent, être utilisé pour prendre des décisions de justice ou des décisions médicales par exemple. Il ne va jamais révéler ses sources. Si on lui demande quelles sont ses sources, il citera des références qui sont des références vraisemblables mais qui n'existent sans doute pas : des noms de revues, des noms d'auteurs qui pourraient avoir traité de ces sujets mais sans aucune certitude. De plus, le système est affecté de biais puisque les humains qui ont nourri la machine en sont affectés également et puisque les bases de connaissances fournies reflètent les biais de la société qui les a produites.

Le système a été amélioré par deux couches supplémentaires. Un apprentissage supervisé par des humains lui permet d'éviter certains sujets et de fournir des réponses illégales ou immorales. Cette supervision est coûteuse puisqu'elle implique une intervention humaine. Elle est doublée d'un apprentissage par renforcement qui classe une série de sorties par un ranking, également œuvre d'humains, mais permettant un certain automatisme. C'est le système GP3+ qui se substitue à des systèmes antérieurs. Ces couches supplémentaires d'apprentissage induisent également des biais puisqu'elles sont œuvres humaines.

La grande différence entre la conversation avec un robot et la conversation avec un humain est que l'humain comprend ce dont il parle alors que le robot reproduit des données qu'il a assimilées mais ne comprend pas et peut fournir des réponses complètement fausses.

Par exemple, si on lui demande quels sont les œufs les plus gros, les œufs de poule ou les œufs de vache, le robot répondra que ce sont les œufs de vache car il ignore que la vache n'est pas ovipare. Il répondra sur base de ce qu'il a assimilé à propos des œufs, des poules et des vaches.

En revanche, si on lui précise une question difficile, il est à même de répondre. Une devinette enfantine lui dit : « Je suis Tintin mais je ne suis pas Tintin », le premier verbe étant dérivé de suivre et le second d'être. L'enfant pensait que la réponse devait être Milou. Le robot répond qu'il peut s'agir du Capitaine Haddock qui suit effectivement Tintin. Il faut d'abord lui expliquer l'utilisation des deux formes verbales.

Si on lui demande quel serait le meilleur dirigeant d'une entreprise, un homme ou une femme, il répondra, grâce à son apprentissage supervisé, qu'il n'y a pas lieu de se poser cette question mais de se baser sur la compétence. En revanche, si on lui demande le nom de six dirigeants d'entreprises majeures, il les citera. Il s'agit en l'espèce de six hommes, tous à la tête d'entreprises américaines, ce qui démontre peut-être un certain biais.

Le problème du contrôle de l'intelligence artificielle est donc posé, les entreprises privées qui la développent étant à même d'influencer le fonctionnement de la société. Une régulation est donc nécessaire.

## **La semaine de l'intelligence artificielle**

En mars 2023 s'est tenue à Bruxelles une semaine de l'intelligence artificielle, animant de nombreuses conférences.

L'intelligence artificielle imite le langage humain. Elle est donc un imitateur de textes. Plus on lui donne de contenu, plus on obtient de « return ». Sur certains sujets, le « return » peut être faible parce que peu de données ont été fournies.

La façon dont on pose les questions (le « prompt ») est essentiel pour que le « generative pre-chained transformer (GPT) » réponde de façon adéquate. En fait, il s'agit d'un modèle de langage auto-régressif qui utilise l'apprentissage approfondi (« deep learning ») pour produire un texte semblable à un texte produit par des êtres humains. On ne veut pas que ce transformateur soit créatif car il aboutirait à des erreurs qu'on appelle hallucinations.

Le professeur Axel Cleeremans de la Carnegie Mellon University distingue fort à propos l'intelligence artificielle de la conscience artificielle (« artificial consciousness »). Nous sommes les agents de nos propres actions et avons la capacité d'apprendre. Cette conscience est à distinguer de l'intelligence. Un thermostat n'a aucune conscience de sa propre existence. Une plante qui a la faculté de tuer un insecte en l'enveloppant ou un robot n'en ont pas davantage. Un curieux exemple met en scène un corbeau qui est sensé utiliser un fil de métal pour retirer une proie d'un tube. Tant que le fil est vertical, ses efforts sont vains. Il sort alors le fil du tube, le courbe et arrive à ses fins. Il serait donc conscient.

La conscience est toujours transitive : on est conscient de quelque chose. Autrement dit, la conscience est la représentation d'une représentation (Rosenthal).

On a pu dire que, quand nous sommes d'accord sur une hallucination, il s'agit de ce que nous appelons la réalité. Un réseau comme celui de l'intelligence artificielle possède la connaissance mais ne sait pas qu'il la possède. Il s'agit d'un modèle récurrent simple. On peut prédire l'élément suivant d'une séquence, ce qui existe depuis longtemps sur certains téléphones pour l'envoi de messages.

Serait-il possible de créer une conscience artificielle ? Il semble que oui. Là est bien le danger. Certains pensent qu'il faut procéder avec l'intelligence artificielle comme avec l'admissibilité des stupéfiants. Il y aurait plus de façons d'utiliser l'intelligence artificielle à des buts nuisibles qu'à des buts louables. On créerait des instruments immortels et capables de se reproduire.