

## «Les machines n'ont pas d'intelligence. Mais elles en auront»



Selon Jeff Hawkins, notre néocortex crée en permanence des milliers de modèles du monde, qu'il met en commun. L'intelligence artificielle s'en inspire.

Le chercheur Jeff Hawkins explore le fonctionnement du néocortex. Passionnant.

Gregory Wicky

Parcours peu banal que celui de Jeff Hawkins. En tant qu'ingénieur en informatique, il fonde dans les années 1990 l'entreprise Palm, qui proposa au monde ses premiers ordinateurs de poche. Désormais neuroscientifique, il dirige une équipe qui tente de modéliser le cerveau humain, notamment dans le but de développer des intelligences artificielles (IA) qui s'en inspirent.

Dans son nouveau livre, «1000 cerveaux», le chercheur explore en détail le fonctionnement de notre néocortex. Partie «récente» du cerveau occupant environ 75% de son espace, il est responsable de toutes les fonctions cognitives «supérieures», comme le raisonnement spatial ou le langage. On le distingue des régions anciennes de l'appareil cérébral, présentes également chez les reptiles ou les amphibiens, sur lesquelles il s'est greffé au fil de l'évolution.

Vulgarisé juste ce qu'il faut pour des lecteurs peu versés dans les sciences, l'ouvrage développe la théorie - saluée en préface par le célèbre biologiste Richard Dawkins pour sa fulgurance digne de Darwin - selon laquelle notre néocortex fait sens du monde en en générant non-stop des milliers de modèles. Hawkins expose ensuite les pistes qui permettront à l'intelligence artificielle de l'imiter, et ainsi faire progresser l'humanité - nous avons affaire à un optimiste. Le chercheur nous répond par Zoom depuis la Silicon Valley.

### **Pourriez-vous résumer simplement la théorie des «1000 cerveaux» ?**

Disons que pour appréhender le monde dans lequel nous vivons, notre néocortex construit un modèle permettant de recréer dans notre esprit tout ce que nous savons. Un peu comme un architecte ferait une maquette de son projet, mais en infiniment plus complexe: tous les gens et les endroits que nous connaissons, tout ce que nous avons appris. Deux facteurs sont déterminants pour nous permettre de créer ce modèle. D'abord, nous le faisons par le mouvement: de nos mains sur des objets, de notre corps dans l'espace, de nos yeux. C'est comme ça que nous apprenons. Le deuxième facteur est que nous ne créons en réalité pas un seul modèle, mais des milliers, que nous mobilisons en permanence.

## **Comment cela fonctionne-t-il?**

Le cortex est structuré par quelque 150'000 colonnes, qui font environ la taille d'un grain de riz. Les neurones y sont particulièrement nombreux et densément connectés. Chacune de ces colonnes est une machine à créer des modèles, qu'on appelle également des référentiels. C'est une construction mentale multiple, distributive, et pas monolithique, comme on l'a longtemps pensé. Les colonnes, présentes à l'identique dans toutes les régions du cortex, créent leur propre modèle qu'elles mettent en commun, puis «votent» sur la perception à adopter et la réponse à apporter.

## **Vous avancez que nous fabriquerons un jour des intelligences artificielles fonctionnant selon ce principe?**

Oui. Ces référentiels créent comme une sorte de cartographie mobilisant à la fois les représentations spatiales, les connaissances, les souvenirs. Chez Numenta (ndlr: l'entreprise qu'il dirige), nous réalisons déjà des versions rudimentaires de tels référentiels. À terme, les IA seront capables d'apprendre par elles-mêmes, de s'ajuster toutes seules à leur environnement, sans qu'on doive investir des millions de dollars et des semaines entières à les abreuver de données.

## **Vous dites qu'il n'y a pour l'heure pas d'intelligence dans l'IA?**

Les nouveaux systèmes dont on parle beaucoup aujourd'hui, comme ChatGPT ou DALLÉ, sont très bons, et continueront de se développer car ils ont un gros potentiel commercial. Mais ils sont fragiles: ils font des erreurs, on les trompe facilement. Ils utilisent une quantité incroyable de données pour tenter d'imiter ce que ferait un être humain en répondant à des requêtes. Mais c'est un tour de passe-passe. Il n'y a pas d'ancrage dans le monde, pas de réflexion. La plupart des chercheurs dans le domaine s'accordent à le dire: à terme, l'IA sera tout autre chose.

## **Vous prenez l'exemple de robots chargés de créer sur Mars des environnements capables d'accueillir des êtres humains**

C'est un exemple un peu fantaisiste mais pas non plus irréaliste. Pour collaborer, surmonter des obstacles, apprendre de leurs erreurs, de telles machines devraient bénéficier d'intelligence artificielle générale, c'est-à-dire pas destinées à une seule fonction, comme la plupart des IA actuelles. Pour l'heure, un enfant de 5 ans qui observe le monde et voit comment une porte s'ouvre ou un verre d'eau se renverse est bien mieux équipé que n'importe quelle IA. Ces machines devraient également être équipées de senseurs, donc posséder quelque chose comme un corps, qui les ancre dans le monde. Car c'est le mouvement dans l'espace qui permet d'apprendre. Ce ne serait pas forcément un corps de robot, type cyborg de science-fiction, mais il lui faudrait une forme de matérialité.

## **On suppose que vous n'êtes pas de ceux qui pensent, comme Elon Musk, qu'il faut un moratoire sur l'IA.**

Non, en effet. Je pense qu'il faut qu'on fasse attention à cette technologie, car elle est puissante et se développe vite. C'est un peu comme les réseaux sociaux: ils ont eu des conséquences négatives et positives, il faut surveiller, légiférer, poser des garde-fous. Mais là n'est pas mon champ de réflexion. La plupart des gens qui ont signé l'appel dont vous parlez ne comprennent pas très bien comment cette technologie fonctionne. Ils ont peur que soudain nous fassions le progrès de trop, et hop, ce sera la fin du

monde. Il n'y a aucune chance que cela se produise. Quant à Elon Musk, il a certes du talent, mais je crois qu'on lui accorde un peu trop de crédit.

### **Vous êtes plutôt optimiste quant aux menaces que peut représenter cette technologie ?**

Les êtres humains sont complexes. Même si notre néocortex est très développé, nous restons des animaux, avec une partie de cerveau ancienne et des instincts de survie, de reproduction. Nos aspirations plus élevées, nos questionnements philosophiques sont souvent en conflit avec nos pulsions et nos désirs, qui peuvent nous pousser à mentir, à voler, à souhaiter du mal à autrui. La plupart des gens peinent à séparer ces deux choses. Ils pensent que l'une n'est pas possible sans l'autre. Mais le type de machine intelligente que nous voulons créer ne sera pas sujette à ces pulsions. Et elle ne pourra pas les développer. Les IA ne peuvent pas nous poser de risque existentiel, vouloir nous détruire ou prendre le contrôle du monde. Le vrai danger lié à cette technologie, c'est qu'elle soit utilisée à des mauvaises fins: désinformation, abus de pouvoir, etc. Ce risque-là, bien sûr, est réel.

### **Il faudra bien pourtant que ces machines soient dotées d'objectifs**

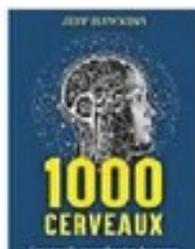
Oui, bien sûr, il faudra leur poser des buts. La voiture automatisée doit vous amener d'un point A à un point B sans percuter d'enfants en chemin! Mais elle ne va jamais développer des objectifs du genre: «J'aimerais avoir une plus grosse maison que mon voisin.» Ça, ça nous est propre. L'IA ne fera rien à moins que vous ne lui donniez quelque chose à faire, c'est facile à implémenter. On peut imaginer quelque chose comme les trois lois de la robotique d'Isaac Asimov (voir encadré).

### **Reste la question de la conscience. Vous pensez que les machines en seront équipées**

Je dis dans le livre qu'elles ne le seront pas forcément, mais que nous pourrons le faire si nous le souhaitons. Le concept de conscience, je pense qu'on a tendance à en faire une trop grande affaire. À mon sens, il se décompose en deux éléments: le sens de sa propre existence, et une forme de mémoire. Ce deuxième élément implique de se souvenir de ses états précédents, ainsi que des choses qu'on a faites jusqu'ici. C'est naturellement indispensable pour apprendre. Les IA pourront être dotées de ces fonctions sans problème.

### **Ce ne sera donc pas un meurtre que de les débrancher ?**

Non. Déjà, nous, humains, nous nous débranchons toutes les nuits au moment du coucher, et nous nous rallumons au réveil. Quant au problème que pourrait poser le fait de détruire une machine consciente, là aussi je n'en vois pas. La peur de la mort nous vient de notre cerveau ancien. Sans lui, pas de crainte ou de chagrin. Les machines s'en fichent bien qu'on les éteigne, qu'on les démonte, ou qu'on les envoie à la casse.



«1000 cerveaux»  
Jeff Hawkins  
Éd. Quanto, 296 p.

# Les trois lois de la robotique selon Isaac Asimov

Les trois lois de la robotique de l'auteur de science-fiction Isaac Asimov furent proposées pour la première fois dans sa nouvelle «Cercle vicieux» en 1942, avant d'être peaufinées par lui-même et d'autres au fil des décennies. Elles sont les suivantes :

1. Un robot ne peut porter atteinte à un être humain ni, restant passif, laisser cet être humain exposé au danger.
2. Un robot doit obéir aux ordres donnés par les êtres humains, sauf si de tels ordres entrent en contradiction avec la première loi.
3. Un robot doit protéger son existence dans la mesure où cette protection n'entre pas en contradiction avec la première ou la deuxième loi.

En 2020, un député français a déposé une proposition de loi pour promulguer une charte de l'intelligence artificielle et des algorithmes. L'article 2 de cette charte est constitué de ces trois lois.

GWY

---

## La théorie des «1000 cerveaux»

### Les 1000 cerveaux de Pascal Kaufmann

D'après lui - et ce qu'il appelle la théorie des mille cerveaux - «le cerveau est en réalité organisé en plusieurs centaines de milliers de petites unités de calcul identiques et indépendantes, les colonnes corticales.»



Plus jeune, il voulait devenir astronaute. Aujourd'hui, il poursuit une autre mission: mettre en place un réseau de super-cerveaux afin de déchiffrer le code neuronal et de créer les bases d'une intelligence artificielle «similaire à celle de l'homme.»

«Enfant, je construisais des robots en Lego. Certes, ils ressemblaient à des humains, mais il leur manquait quelque chose. A l'âge de 12 ans, j'ai lu quelque chose sur Prométhée, le dieu qui avait dérobé le feu et qui avait insufflé une âme à l'Homme. Et je me suis dit: «Quel héros, ce Prométhée!»

A l'époque, son professeur a informé Pascal Kaufmann que seuls les dieux pouvaient créer les hommes. Mais cela ne le préoccupe guère. Il veut être le créateur de l'intelligence artificielle (IA) «similaire à celle de l'homme.» Le scientifique est convaincu que l'IA permettrait d'accomplir des progrès considérables par exemple en médecine. L'IA doit être développée au service de l'homme et la Suisse est prédestinée pour être une «plateforme de talents».

### «Notre cerveau n'est pas un ordinateur»

Selon Pascal Kaufmann, l'IA est l'inverse de la numérisation. Cette dernière se résume à l'exécution continue des mêmes règles. Or pour créer quelque chose de nouveau, il faut enfreindre certaines règles. L'entrepreneur qui souhaiterait penser avec un millier de cerveaux aimerait avoir une petite lentille devant les yeux. Lorsqu'on lui poserait des questions difficiles, les bonnes réponses seraient immédiatement affichées sur celle-ci. Pour créer une telle lentille, on pourrait utiliser la technologie proposée par son entreprise. «Starmind relie des humains à une sorte de superorganisme.» Certes, les évolutions dans le domaine de l'IA peuvent susciter certaines craintes, mais il nous appartient de décider comment utiliser cette technologie.

Aujourd'hui, les hommes se font les mêmes réflexions à propos du cerveau que nos ancêtres lorsqu'ils contemplaient les étoiles en se disant que le mystère qui entoure le déplacement de leurs orbites ne serait jamais percé. Nous sommes subjugués par les milliards de cellules de notre cerveau et pensons que jamais nous ne parviendrons à les comprendre. Selon Pascal Kaufmann, notre cerveau n'est pas un ordinateur, mais un superorganisme composé d'une multitude de petits acteurs. Les lois qui le régissent doivent à présent être déchiffrées par un réseau de super-cerveaux du monde entier. Le neuroscientifique en est convaincu: «si l'on parvient à déchiffrer ce code neuronal, on pourra créer l'intelligence artificielle.»

### «La formation est notre arme la plus précieuse»

D'après Pascal Kaufmann, la formation «est l'arme la plus précieuse» dans un monde automatisé. Nous sommes en mesure de faire la distinction entre les faits, la fiction et la manipulation. Or, dans les écoles, on enseigne encore des choses que les machines maîtrisent déjà mieux que nous. «Nous devrions apprendre à poser les bonnes questions et à faire preuve d'esprit critique», ajoute-t-il.

Le penseur visionnaire estime que ces technologies nous donneront la possibilité de pouvoir davantage décider quelle vie nous voulons vivre. Selon lui, nous avons la chance de vivre dans «une nouvelle époque merveilleuse» dans laquelle nous n'avons peut-être plus besoin de travailler. Ne plus travailler? Mais qu'allons-nous faire de tout ce temps libre? Pascal Kaufmann répond laconiquement: «lorsque l'on travaillait 16 heures par jour dans les champs, la même question s'était certainement posée lors de l'introduction des machines.»

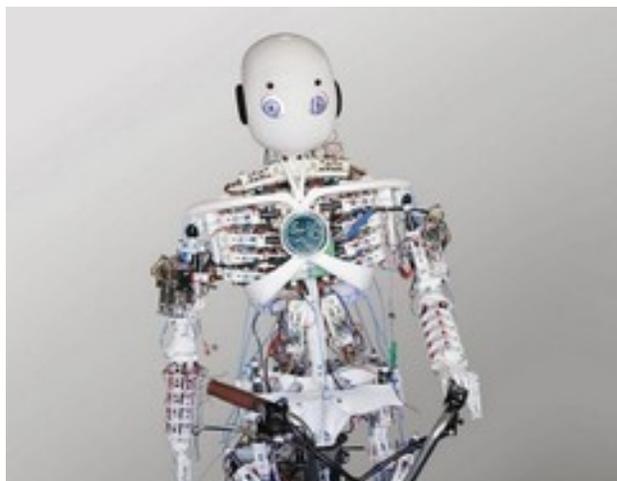
**Texte:** Yvonne Eckert

[www.starmind.com](http://www.starmind.com) | [www.mindfire.global](http://www.mindfire.global)

**Pascal Kaufmann** (né en 1978) a suivi des cours de grec ancien et de latin au lycée et a ensuite étudié les neurosciences. Il a participé au développement de cyborgs au Ministère américain de la Défense. En 2013, il a construit Roboy, un robot humanoïde, au laboratoire d'Intelligence Artificielle de l'université de Zurich. En 2010, avec son ancien camarade d'études Marc Vontobel, il crée Starmind, la société à l'origine de la fondation Mindfire.

## Roboy est un robot humanoïde avancé

Roboy est un robot humanoïde avancé qui a été créé dans le laboratoire d'Intelligence Artificielle (AI-Lab) de l'université de Zurich et qui a été présenté en mars 2013. Initialement, il a été conçu pour aider les personnes au quotidien. Rolf Pfeifer, directeur de l'AI-Lab de l'Université de Zurich, et Pascal Kaufmann, CEO de StarMind International AG, sont les pères spirituels de ce robot.



Publié le 11.04.2019

---

Lettre de lecteur 24heures du jeudi 14 avril 2023

### Technologie

#### **Faut-il vraiment craindre l'intelligence artificielle ?**

**Ainsi, des centaines d'experts, dont Elon Musk, proclament par une lettre ouverte que les systèmes d'intelligence artificielle (IA) «peuvent présenter de grands risques pour la société et l'humanité».**

Selon eux, de tels systèmes «ne devraient être développés qu'une fois que nous sommes convaincus que leurs effets seront positifs et que leurs risques seront gérables». Cela s'apparente au mythe de l'invention d'une langue parfaite qui permettrait uniquement de dire la vérité mais pas de mentir. En pratique, les effets bénéfiques et nocifs d'une invention dépendent beaucoup moins de l'invention elle-même que de l'imagination plus ou moins perverse de ceux qui l'utilisent.

Autre demande des experts: que les systèmes d'IA soient «précis, sûrs, interprétables, transparents, robustes, harmonisés, dignes de confiance et loyaux». Amen! Ils ont oublié d'ajouter «humoristiques»: à mon avis, une IA dépourvue d'humour présenterait de grands risques pour la société et l'humanité. Enfin, les auteurs de cette lettre ouverte appellent «tous les laboratoires d'IA à suspendre immédiatement pendant au moins six mois la formation des systèmes d'IA plus puissants que GPT-4». Tous les laboratoires du monde? Y compris les laboratoires chinois?

Soyons sérieux: si l'Occident s'impose un moratoire, il ouvre un boulevard à une IA communiste et despotique. Sans parler de mouvements mafieux ou terroristes qui peuvent aussi avoir d'excellents

informaticiens parmi leurs membres. D'une façon générale, la tentation d'un moratoire relève de la logique de la prohibition, dont les méfaits catastrophiques sont connus depuis plus d'un siècle.

Pierre Zweiacker, Pully