

> Daniel Andler

# Intelligence artificielle, intelligence humaine : la double énigme

Gallimard, 2023, 432 pages, 25 €

## ▼ L'auteur

Mathématicien, professeur émérite de philosophie à Paris-Sorbonne, spécialiste de la philosophie des sciences, directeur de recherche au CNRS, **Daniel Andler** est actuellement membre de l'Institut universitaire de France et de l'Académie des sciences morales et politiques. Il anime le département d'études cognitives à l'École normale supérieure.

## ● Point de vue

L'auteur livre une remarquable analyse critique de l'ambition prométhéenne de l'intelligence artificielle (IA), qu'il retrace, avec une distanciation (non-jugement, objectivité, impartialité) qui force le respect, à travers l'étude approfondie des réalisations et des échecs des équipes scientifiques, industrielles et techniques depuis trois quarts de siècle.

## ■ L'essentiel

Daniel Andler dit se ranger parmi les sceptiques, pour qui les progrès et les perspectives actuelles de l'IA ne constituent pas une raison de penser qu'elle se dirige vers la création d'une machine intelligente au sens humain. En revanche, l'IA peut mener à l'automatisation d'un grand nombre de fonctions cognitives au bénéfice de chacun.

Ni mystère (qui dépasse l'entendement), ni problème (simple tâche à résoudre), l'IA constitue une énigme à résoudre, au même titre que l'intelligence humaine, les deux étant étroitement liées. Daniel Andler annonce en introduction son intention de montrer que l'objectif de l'IA (créer une intelligence artificielle qui serait l'égale de l'intelligence humaine) est une chimère.

## Le projet, l'entreprise, le parcours

### 1. L'intelligence artificielle survolée

Selon l'auteur, l'espoir d'Alan Turing (« Les machines pourront égaler les hommes dans tous les domaines purement intellectuels », 1950) n'est pas réaliste. Après avoir défini l'IA et précisé le contenu de l'ambition initiale du programme de recherche lancé dans les années 1950, à l'origine des sciences cognitives contemporaines, Daniel Andler recense les cinq étapes du « repli stratégique » de l'entreprise : renoncement à la pensée « pensante »,

à l'équivalence avec l'intelligence humaine, à une intelligence générale, à l'indépendance à l'égard de l'intelligence humaine, à la réflexion.

Verdict de l'auteur : « L'intelligence artificielle [...] n'a pas de rapport direct avec l'intelligence humaine : elle ne comprend pas le monde au sens où les humains le comprennent ; elle n'est pas une qualité unique permettant d'accomplir les tâches dont les humains sont capables. »

### 2. Les sources de l'intelligence artificielle

Après avoir rappelé que l'IA s'appuie sur l'informatique, qui se base sur la logique formelle, dont les calculs s'effectuent *via* des algorithmes, Daniel Andler retrace les trois périodes du projet cybernétique, science de l'information, théorie du contrôle et de l'autonomie des systèmes intelligents, également à l'origine des sciences cognitives. Il analyse les débats sur les promesses, les avancées, les limites et les déceptions à chaque période (classique, connexionniste, actuelle).

### 3. L'ère classique

Première période, à partir du milieu des années 1950 : l'IA symbolique postule alors que l'intelligence (humaine ou artificielle) consiste à soumettre des états ou propositions logiques à des algorithmes ; elle se réduit à la manipulation de symboles selon des règles formelles.

### 4. L'âge du connexionnisme

Deuxième période : prenant acte de l'impasse à laquelle mène l'IA symbolique, le connexionnisme (dans lequel s'inscrit le *deep learning*) s'inspire des modèles neuronaux du cerveau humain. Les systèmes artificiels intelligents (SAI) peuvent alors faire l'apprentissage (supervisé par des humains) de tâches à partir de bases de données, notamment en matière de reconnaissance de formes. En somme, l'IA symbolique est rationaliste, l'IA connexionniste est empirique.

### 5. Les prémices d'une troisième époque ?

Troisième période : l'IA devient générative et se fait connaître du grand public via les modèles massifs de langage (MML), tels GPT (*generative pre-trained transformer*) de la société OpenAI. C'est une véritable révolution qui se produit en 2018 avec le traitement du langage naturel (TLN). L'apprentissage du DL s'effectue désormais aussi par renforcement (behaviorisme) ou de manière autosupervisée. Pourtant, une nouvelle fois, les erreurs parfois grotesques de l'IA générative confortent l'auteur dans son scepticisme quant à la capacité de l'IA à imiter l'intelligence humaine.

Le fait que les SAI soient désincarnés, sans corps, donc sans capacité d'explorer leur environnement, d'apprendre à réagir aux stimuli et par conséquent de devenir autonomes, constitue un sérieux obstacle à ce projet. D'où les espoirs placés dans la robotique, dont l'objectif est de produire des systèmes aptes à réagir à leur environnement, s'adapter à leur milieu et prendre des décisions.

Dernier thème abordé dans la première partie, l'intelligence augmentée dont l'ambition est la démultiplication des capacités de l'esprit humain grâce à Internet et à la miniaturisation des outils informatiques. Comme pour la robotique, l'auteur prend acte des progrès en cours, tout en continuant à s'interroger sur les limites des deux projets.

## La question de l'intelligence et l'avenir de l'IA

### 1. Peut-on comparer intelligence artificielle et intelligence humaine ?

Si les scientifiques, philosophes et industriels définissent l'intelligence comme la capacité à résoudre des problèmes,

comme en sciences cognitives, l'IA peut-elle égaler, voire dépasser, l'intelligence humaine ? L'auteur analyse les controverses entre les optimistes et les pessimistes, les visionnaires et les sceptiques, à partir des arguments des uns et des autres quant à la pertinence de ce qu'il nomme « la Comparaison » (avec un C majuscule) entre les deux types d'intelligence – « Comparaison » qui ne vaut que si l'on définit l'intelligence comme la capacité à résoudre des problèmes, ce qui peut être réducteur... En effet, *quid* de la conscience, du sens (au sens de « sémantique »), du sens commun, de la métacognition et des affects, actuellement inaccessibles aux SAI et pourtant nécessaires à toute démarche (action, pensée, décision) humaine ?

### 2. Intelligence et animalité

L'intelligence doit donc être redéfinie, afin de déterminer si le projet d'origine est réaliste. Daniel Andler propose de faire un détour par l'intelligence animale pour comparer intelligence humaine et IA. Cette démarche permet de contourner l'obstacle de l'intériorité, du flux mental de la conscience qui semble consubstantiel à l'intelligence humaine et introuvable dans l'IA, mais aussi d'insister sur le rapport (biologique) de l'animal (et donc de l'humain) à son milieu (*Umwelt*), à son environnement, qui ne peut en aucun cas être négligé pour comprendre son comportement.

Dès lors, l'intelligence ne peut plus être réduite à la résolution d'un problème dans une situation (par exemple, trouver à manger dans la savane), puisque toute « situation » est le produit conjoint de circonstances et de la « perspective » de l'animal (par exemple, la savane présente diverses ressources alimentaires aux yeux de différentes espèces). D'où la définition proposée par l'auteur : « l'intelligence est la capacité d'agir au mieux dans une situation », définition à la fois descriptive, normative (jugement de fait) et appréciative, subjective (jugement de valeur).

Cette conception de l'intelligence permet enfin d'introduire la notion précieuse de « compréhension » à partir du postulat suivant : « une constellation de croyances et d'informations entrantes qui, à un moment donné, s'imposent à un être humain, voire à un animal non humain, conjuguées à ses désirs et intentions du moment, sollicitent une action de sa part, un mouvement de la pensée qui lui permette de répondre au besoin de "faire ce qu'il faut" » pour atteindre ses fins, survivre, en s'adaptant à son milieu.

### 3. Intelligence et animalité

Pour autant, l'intelligence humaine se distingue de l'intelligence animale par l'autonomie de l'agent capable de « problématiser » une situation et de concevoir un large éventail d'actions, de réactions, de comportements ; l'homme est tout de même moins déterminé par son envi-

ronnement que l'animal. Cette autonomie justifie l'analyse éthique de toute situation/problème (particulièrement complexe chez les humains), puisque, comme écrit plus haut, la capacité d'« agir au mieux dans une situation » relève à la fois d'un jugement de fait et d'un jugement de valeur.

#### 4. L'intelligence artificielle rêvée

En conséquence, « l'IA réussira à coup sûr à automatiser une part croissante des opérations cognitives accomplies par l'homme pour résoudre divers problèmes », mais « la quête d'une IA "forte" ou "prométhéenne", à l'image de l'intelligence humaine et pouvant s'y substituer, est vaine ».

Prenant acte de cette impasse, le courant de recherches nommé IGA (intelligence générale artificielle) vise à concevoir une IA supérieure à l'intelligence humaine, mais cette fois-ci de nature différente : aussi fluide (spontanée et autonome dans la durée) et versatile (capable de résoudre des problèmes de natures diverses), mais fonctionnant différemment et plus puissante. C'est notamment l'ambition des tenants du transhumanisme et de la Singularité. Mais, là encore, l'auteur doute que des SAI deviennent un jour autonomes : versatilité et fluidité ne seront jamais à la portée des « super-intelligences ».

#### 5. L'intelligence artificielle et le bien

Comme en matière de progrès médical, les scientifiques, philosophes et industriels de l'IA ont soumis leur démarche à une réflexion éthique sur ses fins et ses moyens. En effet, l'IA pourrait apporter autant de dommages que de bienfaits à l'humanité. Des centaines de chartes ont été rédigées, définissant les valeurs de la démarche (transparence, justice, équité, innocuité ou non-malfaisance, responsabilité, protection de la vie privée, etc.), mais de grands principes ne suffisent jamais à guider l'action, quel que soit le contexte. En outre, nul ne contrôle totalement une technologie ; par conséquent, ses effets peuvent être contraires aux intentions de ses inventeurs. L'« alignement des valeurs » de l'IA sur celles de l'humain n'est jamais garanti et ne peut l'être. C'est pourquoi, en conclusion de son ouvrage, Daniel Andler préconise le principe suivant : « ne faire appel à l'IA que lorsque sa contribution nette est positive », « lorsque les risques sont réduits et les bénéfices importants ».

■ Emmanuel Peignard