

# Chapitre 13

## L'intelligence artificielle (IA)

Source  
Futura Tech  
Louis Neveu  
2019 ?

### 13.1. L'intelligence artificielle : six usages au quotidien

Dans la culture populaire, l'intelligence artificielle a longtemps été synonyme de robots autonomes qui, soit dans la version catastrophe, viendraient anéantir l'être humain, soit, dans la version humaniste, viendrait le soulager des tâches ingrates. Si cette dernière idée futuriste, commence à se concrétiser sous la forme de voitures autonomes ou de robots aspirateurs ou autres tondeuses, l'intelligence artificielle a déjà cependant largement envahi nos vies.

#### Les e-mails

Les messageries électroniques s'appuient largement sur l'intelligence artificielle afin d'optimiser leur fonctionnement et améliorer l'expérience des utilisateurs. Les filtres des courriers indésirables bénéficient grandement de l'IA, qui peut traiter de grandes quantités de messages et détecter très rapidement les nouveaux spams qui inondent les boîtes de réception.

Pour sa messagerie Gmail, Google utilise TensorFlow, une des IA les plus avancées qui parvient à bloquer plus de 99,9 % des indésirables. Le géant du Web ne s'arrête pas là, et propose d'autres fonctions basées sur l'IA. Sa fonction Smart Reply propose des messages courts pour pouvoir répondre à un e-mail en un seul clic. La firme a aussi développé une fonction baptisée Smart Compose, capable de finir les phrases de l'utilisateur.

#### Les réseaux sociaux

L'intelligence artificielle influence largement la manière dont les informations sont présentées, notamment sur les réseaux sociaux. Par exemple, Facebook trie les publications des contacts et des pages suivies, puis les filtre pour mettre en avant celles jugées les plus importantes, tout en masquant complètement les autres. Le site utilise une IA baptisée DeepText pour analyser le contenu des publications qui ne sert pas uniquement à organiser le fil d'actualités, mais permet également de faire intervenir un employé humain s'il détecte des signes suicidaires.

#### Les moteurs de recherche

Les moteurs de recherche pourraient difficilement fonctionner sans intelligence artificielle, étant donné la taille colossale du Web. L'IA est déployée sous forme de robots qui parcourent les sites pour les indexer, mais également pour les classer et déterminer l'ordre de présentation selon les termes de recherche.

Elle permet même de personnaliser les résultats selon le type de recherche : afficher la météo, la distribution d'un film, ou encore effectuer des calculs ou des conversions de devises. Google propose également sa fonction Google Suggest, la recherche prédictive qui s'affiche pendant la saisie de la requête, qui tente de deviner la recherche basée sur son indexation, les recherches fréquentes et l'historique de l'utilisateur.

#### La traduction

Les services de traduction automatique, accessibles *via* des services comme Google Traduction,

ou directement intégrés à des sites comme Facebook, ont beaucoup progressé ces dernières années, au point de produire des textes parfaitement compréhensibles. Cette avancée a été possible notamment grâce aux techniques de *deep learning*, comme avec l'IA de Google baptisée *Neural Machine Translation system* (GNMT).

### Les sites marchands

L'IA occupe également une place de plus en plus importante pour les sites de vente en ligne. Elle suggère de nouveaux articles à consulter en se basant sur l'historique de l'internaute, ainsi que la navigation des autres visiteurs. Les centres de tri, pour la préparation des marchandises, sont également gérés par IA avec l'utilisation grandissante de robots qui, dans un incessant ballet, acheminent les produits d'un poste à l'autre.

### Les applications de navigation

L'intelligence artificielle a modifié nos habitudes de circulation, grâce à des applications de navigation comme Waze ou Google Maps. Ces dernières déterminent l'itinéraire le plus court et estiment l'heure d'arrivée. Elles peuvent même prendre en compte le trafic en temps réel pour modifier automatiquement la trajectoire afin d'éviter les bouchons.

## 13.2. Les lois d'Asimov

Source

Wikipédia

« Trois lois de la robotique »



Isaac Asimov  
(1920-1992)

Les *Trois lois de la robotique*, formulées par l'écrivain de science-fiction **Isaac Asimov**, sont des règles auxquelles tous les robots positroniques qui apparaissent dans ses romans doivent obéir.

Exposées pour la première fois dans sa nouvelle *Cercle vicieux* (*Runaround*, 1942) mais annoncées dans quelques histoires plus anciennes, les lois sont :

1. un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ;
2. un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi ;
3. un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi.

Au cours du cycle des livres sur les robots, une loi zéro, qui prendra une importance considérable, sera instituée par deux robots, R. Giskard Reventlov et R. Daneel Olivaw, dans la nouvelle *Les Robots et l'Empire*. Cette Loi zéro placera ou tentera de placer la sécurité de l'humanité avant celle d'un individu. Cependant, cette loi n'est pas codée au niveau matériel des cerveaux positroniques, à la différence des trois premières, et elle est une loi de type logiciel, puisque « déduite » par le robot R. Giskard Reventlov.

Les trois lois forment un principe d'organisation et un thème unifiant l'œuvre de fiction d'Asimov, apparaissant dans son *Cycle des robots*, et d'autres histoires reliées à celui-ci, comme dans son cycle de *Lucky Starr*, fiction scientifiquement orientée pour jeune adulte. D'autres auteurs travaillant dans l'univers fictif d'Asimov les ont adoptées, et des références (souvent parodiques) apparaissent dans une bonne part de la science-fiction, et dans d'autres genres. Asimov considérait que ses lois devaient être universelles pour les robots. Aussi, assistant à la projection de *2001, l'Odyssée de l'espace*, il quitta avec bruit la salle lorsque l'ordinateur HAL 9000 viola sa première loi en s'attaquant à des humains.

### De la possibilité d'appliquer des lois à une intelligence

Les mouvances technophile et transhumaniste voient parfois les trois lois comme un idéal à venir. Des avancées significatives en matière d'intelligence artificielle seraient nécessaires pour que les robots obtiennent une intelligence. Toutefois, la complexité des robots a augmenté et l'avancée de la technologie suit une courbe exponentielle selon la singularité technologique. L'humanité a donc intérêt à élaborer des directives et des garanties pour leur fonctionnement.

Roger Clarke a écrit des analyses sur les complications dans l'application de ces lois, dans l'hypothèse que les systèmes soient un jour capables de les employer. Il a fait valoir que les lois de la robotique d'Asimov ont été un procédé littéraire très réussi. Peut-être ironiquement, ou peut-être parce que c'était approprié dans le domaine artistique, la somme des récits d'Asimov réfutent eux-mêmes l'affirmation avec laquelle il commence ces récits, c'est pourquoi Clarke conclut : « Il n'est pas possible de limiter de manière fiable le comportement des robots en concevant et en appliquant

un ensemble de règles ».

En 2004, l'Institut Singularity a lancé une campagne Internet appelée 3 Lois dangereuses : 3 *Laws Unsafe* (les 3 lois d'Asimov) pour sensibiliser aux questions de la problématique de l'intelligence artificielle et l'insuffisance des lois d'Asimov en particulier.

À noter que les trois lois ne sont pas applicables pour les robots militaires qui peuvent avoir pour objectif de tuer des personnes humaines. Dans un tel cas, il est proposé que les robots puissent prendre des décisions éthiques.

Source  
Numerama.com  
Alexis Orsini  
1<sup>er</sup> février 2017

### 13.3. Les « 23 principes d'Asilomar » veulent encadrer le développement de l'intelligence artificielle

2000 signataires, dont l'astrophysicien Stephen Hawking et Elon Musk, patron de SpaceX, ont adopté un guide de référence pour un développement éthique de l'intelligence artificielle. Ces « 23 principes d'Asilomar » ne sont pas sans rappeler les 3 lois de la robotique de l'écrivain Isaac Asimov.

« Parvenir à créer des intelligences artificielles serait le plus grand accomplissement de l'histoire humaine. Malheureusement, il pourrait aussi s'agir du dernier, à moins que nous apprenions à éviter les risques. » L'avertissement lancé en 2014 par l'astrophysicien britannique Stephen Hawking était probablement présent à l'esprit des participants au séminaire Beneficial AI organisé par le Future of Life Institute (FLI, institut pour le futur de la vie) à Asilomar, en Californie, début janvier.

Si les initiatives et les collectifs se multiplient ces derniers temps en faveur d'un développement éthique de l'intelligence artificielle (IA), ces réflexions dispersées n'ont pas encore abouti à un guide commun en la matière. Le FLI espère combler ce manque grâce à ses « 23 principes d'Asilomar » adoptés pendant ces jours de réflexion collective. À ce stade, ils ont été signés par 816 scientifiques spécialisés dans l'intelligence artificielle ou la robotique et 1 200 signataires d'horizons variés, dont des personnalités de renom comme Stephen Hawking ou Elon Musk, le patron de SpaceX et Tesla, qui a participé aux débats.

#### Un guide éthique et pratique

Pour établir ce guide à portée universelle, les scientifiques se sont notamment basés sur les précédents travaux et engagements en la matière, des rapports de la Maison Blanche à celui de l'université Stanford en passant par les propositions des industriels comme Apple et Google, membres du collectif Partnership on AI. Ils y ont trouvé des idées directrices, des éléments communs ou divergents, et en ont tiré une vision d'ensemble simplifiée qu'ils ont ensuite soumise aux participants pour validation.

Ces 23 principes ont tous obtenu au moins 90 % d'approbation de la part des spécialistes présents :

1. **Objectif de ces recherches** : Le développement de l'IA ne doit pas servir à créer une intelligence sans contrôle mais une intelligence bénéfique.
2. **Investissements** : Les investissements dans l'IA doivent être soutenus par le financement de recherches visant à s'assurer de son usage bénéfique, qui prend en compte des questions épineuses en matière d'informatique, d'économie, de loi, d'éthique et de sciences sociales. (Quelques exemples : « Comment rendre les futures IA suffisamment solides pour qu'elles fassent ce qu'on leur demande sans dysfonctionnement ou risque d'être piratées ? » ou encore « Comment améliorer notre prospérité grâce à cette automatisation tout en maintenant les effectifs humains ? »)
3. **Relations entre les scientifiques et les législateurs** : Un échange constructif entre les développeurs d'IA et les législateurs est souhaitable.
4. **Esprit de la recherche** : Un esprit de coopération, de confiance et de transparence devrait être entretenu entre les chercheurs et les scientifiques en charge de l'IA.
5. **Éviter une course** : Les équipes qui travaillent sur les IA sont encouragées à coopérer pour éviter des raccourcis en matière de standards de sécurité.
6. **Sécurité** : Les IA devraient être sécurisées tout au long de leur existence, une caractéristique vérifiable et applicable.
7. **Transparence en cas de problème** : Dans le cas d'une blessure provoquée par une IA, il est nécessaire d'en trouver la cause.
8. **Transparence judiciaire** : Toute implication d'un système autonome dans une décision judiciaire devrait être accompagnée d'une explication satisfaisante contrôlable par un

humain.

9. **Responsabilité** : Les concepteurs et les constructeurs d'IA avancées sont les premiers concernés par les conséquences morales de leurs utilisations, détournements et agissements. Ils doivent donc assumer la charge de les influencer.
10. **Concordance de valeurs** : Les IA autonomes devraient être conçues de façon à ce que leurs objectifs et leur comportement s'avèrent conformes aux valeurs humaines.
11. **Valeurs humaines** : Les IA doivent être conçues et fonctionner en accord avec les idéaux de la dignité, des droits et des libertés de l'homme, ainsi que de la diversité culturelle.
12. **Données personnelles** : Chacun devrait avoir le droit d'accéder et de gérer les données le concernant au vu de la capacité des IA à analyser et utiliser ces données.
13. **Liberté et vie privée** : L'utilisation d'IA en matière de données personnelles ne doit pas rogner sur les libertés réelles ou perçue des citoyens.
14. **Bénéfice collectif** : Les IA devraient bénéficier au plus de gens possible et les valoriser.
15. **Prospérité partagée** : La prospérité économique permise par les IA devrait être partagée au plus grand nombre, pour le bien de l'humanité.
16. **Contrôle humain** : Les humains devraient pouvoir choisir comment et s'ils veulent reléguer des décisions de leur choix aux AI.
17. **Anti-renversement** : Le pouvoir obtenu en contrôlant des IA très avancées devrait être soumis au respect et à l'amélioration des processus civiques dont dépend le bien-être de la société plutôt qu'à leur détournement.
18. **Course aux IA d'armement** : Une course aux IA d'armement mortelles est à éviter.
19. **Avertissement sur les capacités** : En l'absence de consensus sur le sujet, il est recommandé d'éviter les hypothèses au sujet des capacités maximum des futures IA.
20. **Importance** : Les IA avancées pourraient entraîner un changement drastique dans l'histoire de la vie sur Terre, et doit donc être gérée avec un soin et des moyens considérables.
21. **Risques** : Les risques causés par les IA, particulièrement les catastrophiques ou existentiels, sont sujets à des efforts de préparation et d'atténuation adaptés à leur impact supposé.
22. **Auto-développement infini** : Les IA conçues pour s'auto-développer à l'infini ou s'auto-reproduire, au risque de devenir très nombreuses ou très avancées rapidement, doivent faire l'objet d'un contrôle de sécurité rigoureux.
23. **Bien commun** : Les intelligences surdéveloppées devraient seulement être développées pour contribuer à des idéaux éthiques partagés par le plus grand nombre et pour le bien de l'humanité plutôt que pour un État ou une entreprise.

#### **Dans la lignée des trois lois de la robotique d'Asimov**

Ces 23 principes ne sont pas sans rappeler les fameuses 3 lois de la robotique créées par l'écrivain Isaac Asimov, maître de la science-fiction, dans les années 1940. Chacune répond à la précédente : « un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ; un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi ; un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi. »

Reste à savoir si ces 23 principes seront suivis par les concepteurs concernés afin d'éviter les scénarios catastrophe liés à l'émergence des intelligences artificielles, un thème particulièrement porteur dans la fiction.

Source

*Le blog du modérateur*

Thomas Coëffé

13 septembre 2019

### **13.4. Arrêtons de parler d'intelligence artificielle : cela n'existe pas**

*Les machines, aussi avancées soient-elles, n'ont pas grand chose à voir avec l'intelligence.*

Lors de l'événement GEN 2019 dont le BDM est partenaire, nous avons pu assister à l'allocation de Luc Julia sur l'IA. C'est un sujet qui le passionne : après avoir inventé Siri, l'assistant d'Apple, le vice-président mondial de l'innovation chez Samsung a publié un ouvrage intitulé « L'intelligence artificielle n'existe pas ». Il avance des arguments convaincants qui remettent en cause la notion même d'intelligence artificielle.

### **De la désillusion des années 50 à la victoire contre Garry Kasparov**

L'intelligence artificielle, ou tout du moins ce qu'on désigne vulgairement comme de l'intelligence artificielle, n'est pas un phénomène nouveau. L'IA est apparue en 1956 : on a commencé à modéliser un neurone, estimant qu'on pourrait ensuite simuler un réseau de neurones, le tout étant capable de fonctionner comme un cerveau humain. On pensait alors être en mesure de créer une intelligence artificielle. Cela n'a pas fonctionné : les scientifiques se sont tout de suite attaqués à la compréhension du langage naturel (un domaine très complexe, encore aujourd'hui) et les résultats n'ont pas été au rendez-vous. On est alors entrés dans « l'hiver de l'IA » : les financements ont diminué, les investisseurs ont détourné le regard pendant quelques décennies.

Dans les années 90, les réseaux de neurones sont revenus à la mode avec le machine learning. Ces modèles fonctionnent sur des données, contrairement aux systèmes experts qui se basent sur des règles prédéfinies. La puissance de calcul s'est démultipliée, permettant aux programmes de reproduire des schémas basés sur des milliers d'exemples passés. Ces avancées ont abouti, en 1997, à la victoire d'une machine contre Kasparov aux échecs. On estime qu'il existe 10 puissance 50 coups aux échecs, et il « suffit » de trouver le chemin à partir d'une position pour gagner. Un humain n'est pas en mesure de modéliser autant de coups, Kasparov a logiquement perdu la partie. C'était une première médiatique, mais ça n'avait (déjà) rien à voir avec de l'intelligence.

### **La consommation d'électricité et de données en question**

Dans les années 2000, Internet s'est démocratisé. Conséquence directe : on a pu accéder à de très nombreuses données, un carburant nécessaire à l'entraînement des dites intelligences artificielles. Sur Internet, il y a beaucoup d'images de chats et ces photos sont généralement annotées (« c'est mon chat »). Ces associations permettent d'entraîner les IA : des algorithmes de deep learning ont ainsi pu être validés avec ces images de chats et leurs légendes. Les machines étaient capables de « reconnaître » un chat avec 98 % d'acuité, grâce à l'analyse d'une centaine de milliers de photos de chats.

Pourquoi cela n'a rien à voir avec de l'intelligence ? Parce qu'on estime qu'un enfant a besoin de voir deux chats pour être capable de les reconnaître à vie, avec près de 100 % de réussite, même dans des conditions inhabituelles (de nuit, en peinture...).

Autre exemple : en 2016, une autre machine (Deep Mind) a battu le champion du monde de Go. Le Go est bien plus complexe que les échecs car ce jeu n'est pas mathématiquement fini. On ne connaît donc même pas le nombre de combinaisons (mais on sait qu'elles sont bien plus nombreuses qu'aux échecs). Pour arriver à cet exploit, inimaginable quelques années auparavant, il a fallu construire une machine ultra-puissante, composée de 1200 CPUs, 176 GPUs et plusieurs TPUs (tensor processing units, des puces conçues pour TensorFlow). Résultat : une consommation impressionnante de plus de 400 kWh – soit 20'000 fois plus qu'un cerveau humain, qui ne consomme que 20 Wh. Toute cette débauche d'énergie a été concentrée sur un seul objectif : maîtriser le jeu de Go. Cette machine ne sait rien faire d'autre ; le champion du monde de Go, lui, fait d'autres choses dans sa vie. Des arguments forts qui prouvent une fois de plus que ces algorithmes n'ont rien à voir avec de l'intelligence.

### **Les ratés et les limites de l'intelligence artificielle**

Comparé à l'homme, la machine a d'autres limites et consomme une énergie démesurée pour effectuer des tâches plus ou moins équivalentes. Comme l'homme, elle n'est pas infaillible, mais cette imperfection ne correspond en rien à quelque humanité. La machine est imparfaite car ce sont des hommes qui l'ont conçue. On peut évidemment évoquer la fameuse histoire de Tay, le bot Twitter de Microsoft lancé en 2016 – et débranché 16 heures après. En seulement quelques heures, ce bot était devenu l'un des comptes les plus racistes et les plus misogynes du réseau social. Pourquoi ? Parce qu'un robot se base sur la donnée qu'on lui présente : dans ce cas précis, les réponses d'utilisateurs mal intentionnés sur Twitter combinées à un échantillon de base de données, rassemblant des conversations enregistrées dans le sud des USA, dans les années 50. Ces données biaisées ont abouti à la création de ce bot beaucoup trop énervé. Mais rien d'explicable n'est survenu : le programme a simplement apporté des réponses adaptées aux données précédemment analysées.

D'autres ratés peuvent être évoqués. Il y a quelques années, on nous annonçait que d'ici peu, les voitures autonomes rendraient obsolètes nos bonnes vieilles voitures. Les projections sont désormais moins enthousiastes. Luc Julia estime que « la voiture complètement autonome, capable de toujours

prendre la bonne décision et tout faire toute seule, n'existera jamais ». Il propose de visualiser la place de l'Étoile à 18h, en ce vendredi de mouvement social à la RATP. À cette heure-ci, ce n'est plus de la conduite, c'est de la négociation. Il y a tellement de paramètres à prendre en compte qu'on risque rapidement de se retrouver bloqués dans nos voitures autonomes, qui n'oseront pas prendre les mêmes risques que nous. L'autre exemple qui réduit notre optimisme est directement transmis par Waymo, la filiale d'Alphabet dédiée aux voitures autonomes. C'est sans doute la société la plus avancée sur ce sujet, avec plus de 11 millions de miles parcourus. Mais même son PDG le reconnaît : la voiture autonome n'est pas prête pour nos routes. Les vidéos des voitures de Waymo qui parcourent les rues californiennes sont disponibles sur YouTube. On y trouve quelques pépites, comme cette voiture qui n'arrête pas de s'arrêter... car un piéton transporte un panneau Stop sur le trottoir. C'est un événement inhabituel, la machine ne l'a pas bien interprété car elle n'avait jamais été confrontée à cette situation. Depuis, les ingénieurs ont sans doute intégré cette possibilité dans les systèmes. Mais cet exemple montre que la machine ne s'adapte pas, contrairement à l'humain : seul ce dernier est en mesure de réagir lorsqu'il est confronté à une situation singulière, pour la première fois.

### L'IA, de l'intelligence artificielle à l'intelligence augmentée

Luc Julia n'aime pas le terme d'intelligence artificielle et nous ne pouvons que lui donner raison sur ce point : les systèmes basés sur des données et des algorithmes de machine learning ou deep learning n'ont pas de conscience, ils font simplement des choix parfaitement logiques qui dépendent des données et des règles édictées par leurs concepteurs. Il préfère le terme d'intelligence augmentée : ces systèmes augmentent nos capacités, à nous les humains. Ils nous permettent de mieux faire, faire plus rapidement ou plus facilement, ce que nous faisons déjà avant. L'IA n'est pas magique, c'est l'homme qui la conçoit, qui la contrôle et qui décide.

L'homme est relativement moyen dans toutes ses actions ; alors qu'une intelligence artificielle ne peut qu'être experte d'un domaine très particulier (le jeu de Go, les échecs, la conduite...). Elle n'est pas multifonctionnelle. Elle exécute des tâches très précises, qui correspondent à des algorithmes écrits par des humains et un référentiel de données limité. Dans quelques années, il est vrai qu'il sera préférable d'être ramené au domicile par une voiture autonome plutôt que de conduire sous l'emprise de l'alcool. La machine peut être meilleure que l'humain, mais toujours dans des cas très précis.

Et surtout, la machine n'est pas en mesure de s'adapter au futur. Lorsque de nouveaux éléments seront inventés, lorsque notre écosystème évoluera, l'homme s'adaptera. La machine, elle, ne pourra « faire des choix » qu'en fonction de données anciennes qui ne seront pas toujours cohérentes avec cette nouvelle donne. L'IA n'invente pas. Elle peut faire des erreurs lorsque les humains qui écrivent ses algorithmes et lui transmettent des données en font. Les choix des machines sont tous explicables mathématiquement, même si parfois cela prend du temps. On ne peut donc décemment pas parler d'intelligence. Les machines n'ont pas de conscience et ne peuvent pas s'adapter aux situations futures, une faculté propre à l'humanité.

Source  
sciencepost.fr  
Brice Louvet  
28 septembre 2019

## 13.5. Doit-on craindre une prise de contrôle des machines ?

*Alors que l'intelligence artificielle prend de plus en plus de place dans nos vies, certains craignent que cette révolution technologique nous amène à un asservissement des humains par les machines. Mais une IA pourrait-elle réellement dominer notre espèce, de sa propre volonté ?*

Les systèmes d'intelligence artificielle ont intégré nos vies, qu'il s'agisse des domaines de la médecine, de la sécurité informatique, de la conduite autonome ou de la surveillance des foules (et bien d'autres). Le fait est que les machines apprennent vite, deviennent surpuissantes. Alors forcément, une question se pose : une IA pourrait-elle un jour, d'elle-même, prendre le contrôle des humains ? Ce genre de scénario dramatique, les films d'Hollywood nous y ont habitués. Dans *Terminator*, par exemple, un système de défense informatique – Skynet – prend conscience de lui et déclenche une guerre nucléaire, anéantissant une grande partie de l'humanité. Mais est-ce pour autant plausible ? Non, répond Anthony Zador, neuroscientifique au laboratoire de Cold Spring Harbor.

### L'IA non soumise à la sélection naturelle

Cette vision du futur, dit-il, reflète une véritable **incompréhension de ce que représente une intelligence artificielle**. « Nous surestimons considérablement la menace d'une prise de contrôle

*accidentelle de l'IA, car nous avons tendance à associer l'intelligence à la volonté de parvenir à la domination, explique le chercheur. Cette confusion est compréhensible : au cours de notre histoire évolutive, l'intelligence était la clé de la domination sociale. Or l'intelligence en soi ne génère pas de but en soi – comme la volonté de domination. C'est simplement un moyen de l'atteindre ».*

À côté de ça, tous les êtres vivants se caractérisent par leur « intelligence naturelle ». Tous ont, grâce à la sélection naturelle, développé un tas de stratégies visant à perpétuer leur espèce. Cet objectif – rester en vie – est profondément câblé dans le système nerveux de tous les organismes, même les plus simples. Mais les systèmes d'IA n'ont jamais été concernés la sélection naturelle. Ils n'ont ainsi pas eu besoin de développer d'instinct de survie qui mène à l'impulsion de vouloir dominer les autres.

Une intelligence artificielle ne pourra donc pas, d'elle-même, prendre la décision d'asservir une autre espèce dans le simple but de survivre à tout prix. Elle ne pourra le faire que si une entité (une personne, une entreprise) lui demande de le faire.

### **Se concentrer sur les véritables risques**

Pour le chercheur, cette inquiétude vis-à-vis de cette potentielle « prise » de contrôle des machines nous distrait des véritables risques de l'IA. Dans un futur proche, si ce n'est pas déjà fait, certains systèmes intelligents seront transformés en armes autonomes. Quand d'autres, en remplaçant nos emplois, auront le pouvoir de perturber toute une économie en place depuis des décennies. L'accessibilité limitée à ces systèmes pourrait également creuser davantage les inégalités sociales. Sans oublier les débordements imprévus. Tous ces problèmes sont plus inquiétants qu'une prise de contrôle par les robots.

Mais ces risques, néanmoins, peuvent être contenus si nous en prenons rapidement conscience. Et ce qui nous faisait autrefois peur pourrait en réalité être tourné à notre avantage. Des armes plus puissantes et intelligentes pourraient effectivement avoir des effets dissuasifs. L'intelligence artificielle pourrait également stimuler l'économie en étant plus productive, mais aussi nous libérer des emplois les plus contraignants pour en créer de nouveaux, plus accessibles et agréables. Quant aux inégalités, de nouvelles politiques budgétaires peuvent être mises en place.

Au final, il ne fait aucun doute que l'IA va orienter de profondes mutations au cours des prochaines décennies. Mais nous aurons toujours la main dessus. C'est donc à nous de savoir ce que nous voulons (et allons) en faire.