



HAL
open science

Intelligence artificielle : entre science et fiction

Marie Coris, Natacha Vas-Deyres, Nicolas P. Rougier

► **To cite this version:**

Marie Coris, Natacha Vas-Deyres, Nicolas P. Rougier. Intelligence artificielle : entre science et fiction. The Conversation France, 2020. hal-04073390

HAL Id: hal-04073390

<https://hal.science/hal-04073390>

Submitted on 18 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

Intelligence artificielle : entre science et fiction

Publié: 20 novembre 2020, 12:00 CET

Marie Coris

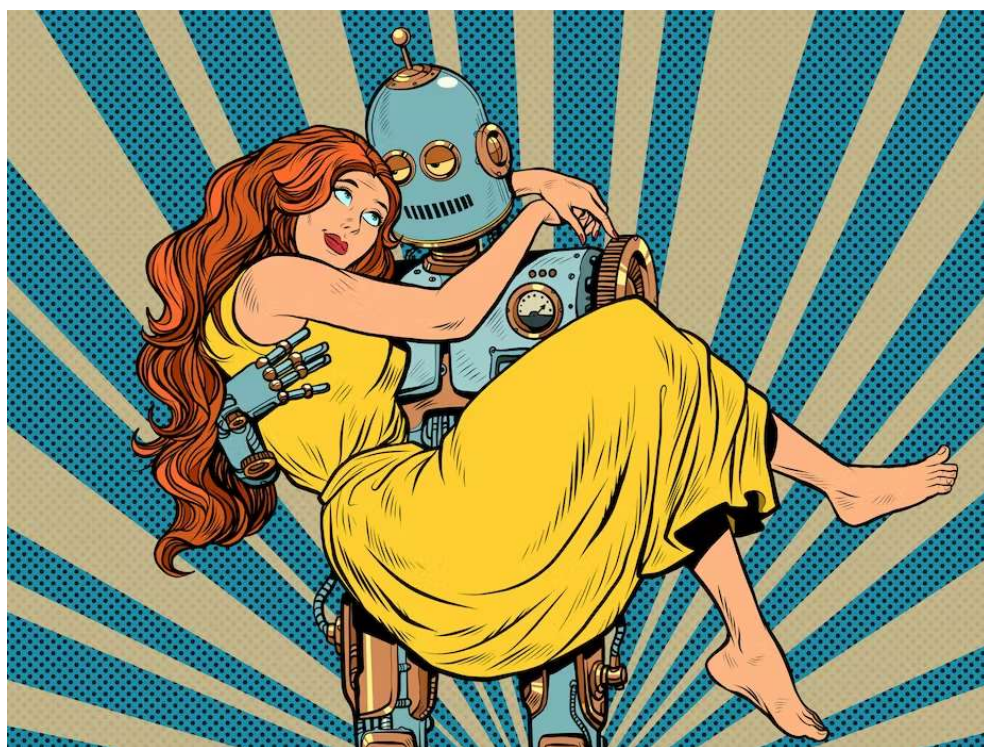
Enseignant-chercheur économie de l'innovation, laboratoire GREThA, Université de Bordeaux

Natacha Vas-Deyres

Chercheuse en littérature française, francophone et comparée du XXe siècle, spécialiste de l'anticipation et de la science-fiction littéraire et cinématographique., Université Bordeaux Montaigne

Nicolas P. Rougier

Chargé de Recherche en neurosciences computationnelles, Université de Bordeaux



L'amour humain, robot : bientôt une réalité ? Studiostoks / Shutterstock

Les fantasmes actuels associés à l'intelligence artificielle (IA) trouvent leur origine dans la science-fiction (SF). Bien avant qu'on parle d'IA, bien avant les découvertes d'Alan Turing dans les années 1950. Bien avant cela donc, les robots – du Tchèque *robota* signifiant corvée – avaient déjà envahi la littérature de SF. En 1883, les robots agricoles ou “*Atmophytes*” de Didier de Chousy dans *Ignis* ou les androïdes de Karel Čapek en 1920 dans sa pièce de théâtre *R.U.R.*, offrent une première représentation du lien étroit entre robots et autonomie de la pensée. Dans *R.U.R.*, les robots finissent par se révolter contre l'humanité. D'emblée dominante, cette représentation destructrice et apocalyptique n'est pourtant pas la seule qui sera véhiculée par la SF.

Entre complexe de Frankenstein et « empathie machinique »

Robots et IA mêlés, la SF a construit un rapport ambivalent aux machines, tout à la fois fondé sur le complexe dit « de Frankenstein » et sur l'empathie machinique. Pour l'écrivain Isaac Asimov, le complexe de Frankenstein, celui de la révolte contre l'humanité, trouve sa source dans l'œuvre de Marie Shelley (1818 !) où la créature artificielle tue son créateur. Un siècle plus tard, un des premiers robots de cinéma, Q l'Automaton (*The Master Mystery*, 1918), renvoie aussi à cette idée.

« Q » The Automaton at The Houdini Museum Of NY at Fantasma Magic/Fantasma Magic.

À l'opposé, l'empathie machinique, c'est l'idée qu'une machine entretient un lien émotionnel avec l'humanité, qu'elle serait prête à tout pour nous protéger (*AI*, l'IA enfant, *Wall-E*, etc). Le premier robot de ce type à voir le jour dans la SF est le personnage d'Adam Link (dans *Amazing Stories* de 1939 à 1942) créé par les frères Binder, Earl et Otto. Isaac Asimov s'en inspirera pour proposer des robots bienveillants envers l'humanité dans des nouvelles consacrées à des héros de métal gouvernés par les trois lois de la robotique (un robot ne peut porter atteinte à un être humain ni, restant passif, permettre qu'un être humain soit exposé au danger ;

un robot doit obéir aux ordres que lui donne 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. Il ira encore plus loin dans « L'Homme bicentenaire » (*Stellar Science fiction*, 1976) : c'est ici le robot qui veut devenir humain pour intégrer notre condition mortelle et abolir son immortalité de machine indestructible.

WALL-E Bande Annonce VF 2008 HD/Films Exclu.

La représentation fictionnelle de l'IA évolue aussi en suivant le contexte des avancées scientifiques. La SF d'aujourd'hui ne reproduit plus, par exemple, cette image d'un ordinateur géant, centralisé et omnipotent mais bien cette IA dématérialisée, présente dans de petites unités, partout et nulle part à la fois, comme dans *HER* de Spike Jonze (2014) ou dans *Les Machines fantômes* d'Olivier Paquet (2019). L'IA est désormais représentée comme une entité impalpable qui envahit le monde. On nage ici en pleine réalité car les IA sont désormais partout (assistant personnel, voiture, téléphone, etc.).

Ainsi, souvent, l'imaginaire dépasse la réalité car il repose sur un postulat indépassable, celui d'une IA ayant surpassé l'humain.

L'IA, une intelligence parmi d'autres

On peut définir l'intelligence comme la capacité à utiliser une expérience passée pour s'adapter à une situation nouvelle. Prenons l'exemple d'Alfred. Si Alfred fait x fois la même erreur face à la même situation, on perçoit le problème. Si par contre Alfred est capable de s'adapter rapidement à un changement dans son environnement pour réaliser une tâche en utilisant des ressources, non proposées immédiatement, mais qui font appel à ce qu'il a appris ailleurs, là Alfred fait preuve d'intelligence.

Les IA sont une sorte d'intelligence parmi d'autres et, comme toutes les intelligences, elles font des erreurs pour trouver la bonne solution. L'auto-correction et l'amélioration continue nous permettent ainsi d'évoluer tout en apprenant de nos erreurs. Même combat pour les IA, avec l'apprentissage automatique (machine learning) permis par l'accès à un nombre massif de données depuis les années 2010. Plus précisément, la machine cherche des liens entre les données récoltées pour les catégoriser. Elle fait alors preuve d'« intelligence », à l'instar d'Alfred. Mais elle n'est pas autonome de l'action humaine, pour avoir accès aux données mais surtout pour réagir face « à un problème nouveau ».

Prenons le cas d'une voiture autonome. Pour s'adapter à la route, ces véhicules "lisent" le paysage, les panneaux de signalisation, etc. Il suffit cependant d'un petit détail pour tromper l'IA, par exemple d'un petit autocollant posé sur un panneau de signalisation. Nos IA ne sont pas encore assez performantes pour éviter ce type de leurre, d'où la nécessité d'un conducteur humain pour parer à ce type d'éventualité.

Intelligence générale et intelligence spécialisée

La croyance dans les représentations fictionnelles de l'IA provient aussi du fait que le grand public ne fait pas la distinction entre deux sortes d'intelligences, celle considérée comme « générale » et celle dite « spécialisée ». On parle d'intelligence générale pour désigner une capacité d'apprentissage et d'adaptation quasi infinie et très rapide. Elle permet de prendre des décisions tout en se plaçant dans un contexte moral. L'intelligence humaine, ou animale, est une intelligence générale. L'intelligence spécialisée désigne, elle, la capacité à agir sur des tâches spécifiques ou des objectifs extrêmement précis. C'est là où nous sommes aujourd'hui en matière d'IA.

Les IA sont des systèmes qui sont entraînés à réaliser des tâches spécifiques de plus en plus compliquées. Ces systèmes nous semblent alors intelligents. Certains algorithmes peuvent effectivement devenir experts dans un domaine très précis (reconnaissance faciale, jeu d'échecs, etc) mais ils ne savent faire que cela. Votre GPS, par exemple, ne pourra jamais comprendre des images du jour au lendemain, la reconnaissance faciale ne pourra pas planifier votre chemin... enfin pas sans qu'on l'ait explicitement programmée.

L'IA aujourd'hui n'a pas la capacité d'adaptation requise pour prendre des décisions autonomes. Que penser de la situation imaginée par A. Proyas en 2004 dans son film *I-Robot* où le robot « Adam » préfère sauver le policier plutôt que la petite fille car il avait une meilleure capacité de survie ?

L'IA peut-elle devenir plus intelligente que l'humain ?

Des IA plus intelligentes que l'humain ordinaire, est-ce simplement possible ? Surpasser l'humain sur un plan intellectuel sous-entend que l'IA serait capable de prendre des décisions pour nous.

Évidemment que la thématique est source d'inspiration pour les œuvres sciences-fictionnelles qui abordent ici une question essentielle sur laquelle la science n'a pas encore de réponse : quels buts poursuivent les machines ? Le thème de l'annihilation de l'humanité pour créer une civilisation machinique se retrouve dans Matrix (où les machines asservissent les humains à leur insu pour utiliser la chaleur et l'activité électrique de leur corps comme source d'énergie). Wall-E propose une lecture plus empathique de la prise de décision par l'IA. Dès les années 1980, Tron livrait cette vision de la création d'un monde parfait, épuré des erreurs humaines, sorte de première utopie cyberpunk pour machines et programmes mais un enfer pour l'humain.

Les avis des experts sur la question convergent tout de même sur un point : « ce n'est pas pour aujourd'hui ! » Et pour demain ? En l'état actuel des connaissances, rien n'est moins sûr. Le défaut de capacité d'adaptation fait qu'il est possible pour une IA de nous être « supérieure », mais seulement dans des domaines spécialisés (comme dans l'exemple précité du jeu d'échecs ou du jeu de go).

Force est alors de tordre le cou à un autre mythe, celui qui prédit la possibilité d'un grand remplacement de l'humain par une super-IA. Cet événement surviendrait après ce qu'on appelle la Singularité, un moment unique d'une évolution autonome et accélérée où l'IA surpasserait l'humain en intelligence et prendrait en main son propre destin ainsi que le nôtre. Certains vont jusqu'à dire que cela pourrait conduire à la mise en esclavage ou, pire, à l'extinction de l'humanité ! Le problème de la Singularité est qu'elle se heurte à la même limite que l'intelligence générale. Or, cette « thèse » déborde le domaine de l'imaginaire pour nourrir de fausses croyances qui s'ancrent dans la réalité.

Faut-il avoir peur de l'IA ?

Ces mythes masquent le véritable risque d'une IA qui est déjà partout présente dans notre quotidien via les smartphones et autres objets « intelligents ». Ils nous assistent dans nos choix en fournissant des propositions fondées sur nos préférences (en ayant accès aux données que nous stockons, parfois sans le savoir).

Nous pensons disposer d'un certain pouvoir de décision mais au final, c'est la machine qui prend les décisions. L'humain se fait, et volontairement donc, progressivement dépouiller de sa faculté critique. Certains y perçoivent une forme de manipulation, que ce soit pour générer plus de profit au service du marché, ou pour prendre le contrôle sur nos vies. L'IA est déjà utilisée non seulement pour prédire des choix et des comportements mais elle est aussi utilisée pour les influencer. « Celui qui deviendra leader en ce domaine sera le maître du monde », nous dit Vladimir Poutine.

Au final, la question que l'on devrait, démocratiquement, se poser et la suivante : pour quelles raisons voulons-nous des IA ?

Parce qu'on sait le faire ? Parce qu'on ne sait pas le faire ? Parce que d'autres le font ? L'IA n'est pas phénomène naturel qui s'imposerait à nous. Les IA sont des outils informatiques comme les autres et elles doivent donc uniquement être conçues en réponse à des besoins explicites et fournir tous les éléments nécessaires à leur compréhension et à leur utilisation. Informatique et IA sont des moyens, pas des finalités. Et ce ne sont pas forcément, partout et toujours, les meilleurs moyens.

Cet article a été co-écrit avec Karen Sobriol, étudiante en Master 1 Médiation des sciences (université de Bordeaux-Université Bordeaux Montaigne), dans le cadre de son stage professionnel effectué au département de recherches en SHS Changes de l'université de Bordeaux.