

INTELLIGENCE ARTIFICIELLE ET ÉTHIQUE

Avoir peur de notre ombre...

Véronique Barthelemy-Pelletier @VPELLETIER

veronique.pelletier@adeli.org

@ADELI_IT

« Je me suis rendu compte que Gödel, Escher et Bach¹ n'étaient que des ombres projetées dans différentes directions par une essence centrale.

J'ai essayé de reconstruire cet objet central, et c'est ce livre². »

Douglas Hofstadter

Résumé : L'Intelligence Artificielle - IA, (*Artificial Intelligence - AI* en anglais) est abstraite pour beaucoup. Quelques définitions, quelques enseignements, quelques personnalités de l'IA, quelques problématiques, quelques livres et quelques organismes sont évoqués dans cet article qui nous amène à réfléchir au rapport entre l'IA et l'éthique.

Mots-clés : Éthique, Intelligence Artificielle, IA, AI, Politique



J'ai fait le choix, étrange, en 1985 de m'inscrire en Maîtrise d'Informatique (Master I) et de choisir les modules Bases de Données, Téléinformatique et Intelligence Artificielle à l'Université Pierre et Marie Curie (à l'époque Paris VI ou Jussieu). Cette université, appelée aujourd'hui UPMC, fait partie du groupe Sorbonne Universités.

Lorsque l'on est étudiant, c'est le nom qui vous enchante et vous inspire. Ce fut, du moins, mon cas.

J'ai eu la chance de côtoyer d'illustres enseignants cette année-là :

Georges Gardarin pour les bases de données, Guy Pujolle pour la téléinformatique et Jean-Louis Laurière pour l'intelligence artificielle.

¹ https://fr.wikipedia.org/wiki/G%C3%B6del,_Escher,_Bach:_Les_Brins_d%27une_Guirlande_%C3%89ternelle

² Gödel, Escher, Bach: an Eternal Golden Braid – 1979 chez Basic Books – traduction française en 1985 chez Dunod : Gödel, Escher, Bach : Les Brins d'une Guirlande Éternelle



INTELLIGENCE ARTIFICIELLE

Qu'est-ce que c'est ?

Selon Wikipédia, « l'intelligence artificielle est une discipline scientifique appartenant au groupe des sciences cognitives et au groupe des sciences informatiques. Elle recherche des méthodes de résolution de problèmes à forte complexité logique ou algorithmique. Par extension elle désigne, dans le langage courant, les dispositifs imitant ou remplaçant l'humain dans certaines mises en œuvre de ses fonctions cognitives.

Ses finalités et son développement suscitent, depuis toujours, de nombreuses interprétations, fantasmes ou inquiétudes s'exprimant tant dans les récits ou films de science-fiction que dans les essais philosophiques.

Historiquement, elle trouve son point de départ avec les travaux d'Alan Turing qui se demande en 1950 si une machine peut « penser ». Le développement récent des technologies informatiques et des techniques algorithmiques comme le **deep learning** et les **réseaux neuronaux** ont permis la réalisation de programmes informatiques surpassant l'homme dans certaines de ses capacités cognitives emblématiques : le jeu d'échecs en 1997, le jeu de go en 2016 et le poker en 2017. Ces succès encouragent les spéculations. Dans les milieux technophiles on verse, en général, dans l'enthousiasme, le mouvement transhumaniste en est la meilleure expression. Au contraire, d'autres s'inquiètent et sont portées par une perplexité, parfois alarmiste, y compris dans la sphère de la haute technologie. Ainsi, des figures réputées, tel Bill Gates PDG de Microsoft, « figure emblématique de la révolution informatique de la fin du XXe siècle », pense qu'il faut rester très prudent quant aux développements futurs de ces technologies qui pourraient devenir liberticides ou dangereuses. »

Quelques rudiments

Mon professeur, Jean-Louis Laurière, décédé en 2005, nous a enseigné la logique, les syllogismes, les systèmes experts, le modus ponens³, le modus tollens⁴, la recherche opérationnelle, il nous a parlé de théorèmes, de grammaires, de la machine de Turing, de décidabilité, de *représentation des connaissances*⁵, de *résolution de problèmes par l'Homme et la machine*⁶.

Il nous a appris à ne pas nous conformer aux raisonnements habituels mais à toujours changer de perspective pour appréhender un problème. Nous avons beaucoup joué aux échecs en essayant d'optimiser, de choisir le meilleur coup à jouer. À l'époque, l'ordinateur ne battait pas le champion du monde d'échec !

Il nous a parlé de parcours d'arborescences, d'algorithme alpha bêta, de cut-off... Ces parcours d'arbres préfiguraient déjà la logique floue⁷.

La logique floue est une extension de la logique classique qui permet la modélisation des données imprécises, ce qui permet de modéliser des intervalles et non des données discrètes, et se rapproche dans une certaine mesure de la flexibilité du raisonnement humain en utilisant la théorie des possibilités.

³ https://fr.wikipedia.org/wiki/Modus_ponens

⁴ https://fr.wikipedia.org/wiki/Modus_tollens

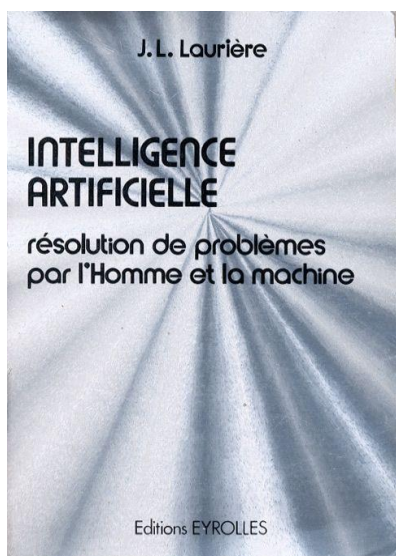
⁵ Jean-Louis Laurière 1988 – INTELLIGENCE ARTIFICIELLE, *résolution de problèmes par l'Homme et la machine*

⁶ Livre de Jean-Louis Laurière 1986

⁷ <https://openclassrooms.com/courses/introduction-a-la-logique-floue>



Jean-Louis Laurière



Jean-Louis Laurière avait une vision différente de la programmation :

« *Penser autrement l'informatique et la résolution de problèmes* ».

Il décrit l'importance de la **modélisation** et des **représentations**, il préconise l'approche :

« *déclarative – Ne plus programmer !* ».

Dans l'avant-propos de son livre, il dit :

« *le pourquoi des choses n'est jamais donné* »

et

« *à quoi cela pouvait-il servir ?* »

et

« *En vérité, ce ne sont que des prétextes pour faire passer quelque chose de plus noble : apprendre à comprendre, apprendre à résoudre des problèmes, apprendre à apprendre* ».

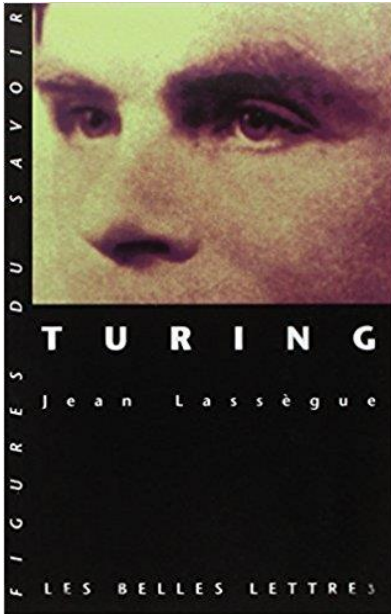
Il était (entre autres) un grand admirateur de Lewis Carroll. Il a d'ailleurs appelé ces programmes SNARK, ALICE et RABBIT.

Systeme expert

Le terme Système Expert n'est plus très utilisé, on parle de systèmes à base de connaissances ou de moteur d'inférences d'ordre 0 (logique des propositions), 0+, 1 (logique des prédicats), 2 avec des variables.

Lorsque l'on a une base de faits et une base de connaissance décrite sous forme de règles, on applique des règles (Si <prémisse> ALORS <Conclusion>), qui permettent de compléter cette base de faits si la prémisse est vérifiée. Le système fonctionne à partir d'un raisonnement mathématique. L'algorithme mis en œuvre est très simple dans les moteurs d'ordre 0. C'est ce que l'on appelle un moteur de règles. C'est une boucle qui vérifie l'applicabilité des règles. Quand la prémisse est applicable, la règle se déclenche et la conclusion vient s'ajouter à la base de faits, ce qui complète la connaissance sur un sujet.

La machine de Turing



Dans son livre *Turing*, Jean Lassègue dans la deuxième partie *la logique du calcul* situe la contribution de Turing en mathématiques théoriques (« théorie de la calculabilité ») au sein du débat mathématique des années trente.

« *La machine de Turing est en réalité virtuelle, « théorique », et son concept est facile à décrire. Un ruban mobile défile devant une tête de lecture. Le ruban est découpé en cases, lesquelles sont soit vides, soit contiennent un symbole. La tête de lecture lit le contenu de chacune des cases et, en fonction d'un ensemble d'instructions qu'elle possède en mémoire, décide de laisser le symbole lu en l'état ou de l'effacer et de le remplacer par un autre. À la suite de quoi la machine déplace le ruban d'une case vers la droite ou vers la gauche. Tout étudiant en informatique découvre avec surprise et amusement que des opérations intuitivement complexes peuvent être effectuées à partir d'un nombre réduit d'instructions du type « si... alors... ». Il est ainsi*

possible de vérifier avec un bout de papier et un crayon qu'une machine de Turing peut très facilement lire deux nombres (selon leur expression binaire faite de zéros et de uns) et écrire leur somme un peu plus loin sur le ruban (ou, le cas échéant, sur un ruban qui défile parallèlement, la tête d'écriture étant alors distincte de la tête de lecture). Ainsi se convainc-t-on sans difficulté que tout calcul, c'est-à-dire toute transformation d'un nombre en un autre en fonction de règles déterminées peut s'effectuer de la même manière. »

Turing fut un des pionniers de l'informatique.

Apprentissage automatique

L'apprentissage automatique⁸ se dit *machine learning* en anglais.

Il permet de doter les machines de systèmes de perception de leur environnement (vision, reconnaissance d'objets, de visages, de textes...).

Il existe un apprentissage supervisé, un apprentissage non supervisé...

L'apprentissage automatique est utilisé pour un large spectre d'applications, par exemple :

- moteur de recherche ;
- aide au diagnostic ;
- détection de données aberrantes ;
- détection de données manquantes ;
- détection de fraudes ;
- analyse des marchés financiers ;
- reconnaissance de la parole ;
- reconnaissance de l'écriture manuscrite ;
- analyse et indexation d'images et de vidéos, en particulier pour la recherche d'image par le contenu ;
- robotique.

⁸ https://fr.wikipedia.org/wiki/Apprentissage_automatique

Reconnaissance des formes

La reconnaissance des formes⁹ correspond à une reconnaissance de motifs. La méthode peut être effectuée à partir d'algorithmes d'apprentissage automatique :

- réseau de neurones ;
- analyse statistique...

Robotique

Les robots d'aujourd'hui sont capables de beaucoup de choses. Chaque robot est indépendant et interagit avec le monde qui l'entoure, avec un autre robot éventuellement. Une voiture autonome est-elle un robot ?

Il perçoit son environnement et apprend par l'expérience, tout comme nous.

Martine Otter a écrit un article sur le droit des robots dans la Lettre n° 103 d'ADELI¹⁰.

Deep learning¹¹ (apprentissage profond)

Cette technologie d'apprentissage est fondée sur les réseaux neuronaux a bouleversé le domaine de l'intelligence artificielle.

Yann LeCun¹² est l'un des inventeurs du deep learning. Il disait en 2015 :

« Je n'ai jamais vu une révolution aussi rapide. On est passé d'un système un peu obscur à un système utilisé par des millions de personnes en seulement deux ans. »

Il est français. Il a rejoint Facebook où il dirige le laboratoire FAIR, Facebook Artificial Intelligence Research qui a ouvert des bureaux à Paris.

ÉTHIQUE

Définition

J'apprécie cette définition trouvée sur internet de l'ordre des ingénieurs du Québec :

« L'éthique est une réflexion sur les valeurs qui orientent et motivent nos actions. »

L'éthique me semble être une façon de se comporter en respectant des valeurs et des principes individuellement et collectivement pour le bien de tous, se dirigeant vers l'intérêt général qui serait un but à atteindre.

Quelques problématiques

Les problématiques soulevées par l'intelligence artificielle sont énormes et imposantes... puisque l'intelligence artificielle est partout aujourd'hui : dans les téléphones portables, dans les salles de marchés, dans les robots, les objets intelligents, la ville intelligente, la médecine, la justice, l'agriculture, la police et l'armée, l'industrie...

La sécurité, l'atteinte à la vie privée, l'atteinte à la dignité de l'homme, la surveillance mondiale et les entreprises qui volent nos données sont des défis à relever.

⁹ https://fr.wikipedia.org/wiki/Reconnaissance_de_formes

¹⁰ <http://www.adeli.org/contenu/lettre-numero-103-printemps-2016-polyvalence>

¹¹ https://fr.wikipedia.org/wiki/Apprentissage_profond

¹² https://fr.wikipedia.org/wiki/Yann_Le_Cun

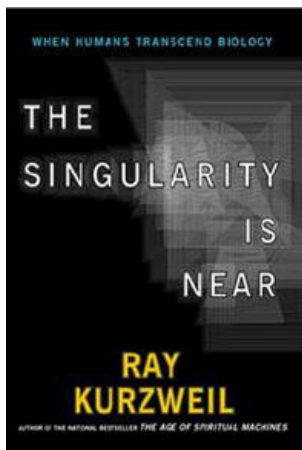
Des machines pourraient faire incarcérer des personnes innocentes...

Discriminer des personnes selon leur sexe, leur couleur de peau serait possible.

Les voitures sans chauffeurs pourraient ne pas « voir » des dangers.

Cyborg et IA - La fusion programmée entre l'homme et la machine¹³ : entre la « mécanisation » de l'homme et l'« humanisation » des machines

Des craintes concernant la perte de contrôle des systèmes IA pourraient menacer l'humanité. Une évolution plus rapide des systèmes vers une « singularité¹⁴ » de l'intelligence puissante. Beaucoup de questions se posent alors pour ne pas y parvenir.



À ce sujet vous pouvez lire le livre de Ray Kurzweil : « The singularity is near ». Il prévoit des impacts sur :

- Le corps humain ;
- Le cerveau humain ;
- La longévité humaine ;
- La guerre ;
- L'apprentissage ;
- Le travail ;
- Les jeux...

Surveillance et vol de données

Surveillance sur le web « visible » :

Foursquare sait qui vous êtes ;

Flickr sait ce que vous voyez ;

Facebook sait ce que vous faites ;

Linkedin sait ce que vous avez fait ;

Twitter sait ce que vous dites ;

Amazon sait ce que vous achetez ;

Google sait ce que vous pensez ;

...

**Courtesy of F. Bancihlon*

¹³ <http://www.commentcamarche.net/contents/2208-cyborg-et-ia-la-fusion-programmee-entre-l-homme-et-la-machine>
¹⁴ https://fr.wikipedia.org/wiki/Singularit%C3%A9_technologique

Sécurité

<http://www.futura-sciences.com/tech/actualites/robotique-intelligence-artificielle-pose-cinq-problemes-securite-selon-google-63293/>

Aide à la décision

L'IA peut servir à résoudre un problème ou aider l'humain dans sa décision. S'il compulse les données et donne un éclairage, pourquoi pas. Mais c'est aussi l'automatisation. Que dire d'une décision prise automatiquement. En théorie, elle est prouvée car l'IA se fonde sur des théorèmes.

Big data

La relation Big Data / IA a été étudiée lors de ce colloque :

<http://bigia2016.irisa.fr/>

Transparence des algorithmes

Les algorithmes sont-ils les nouveaux décideurs ?

Faut-il repenser la responsabilité des acteurs publics ?

Quelle autonomie a l'individu face aux algorithmes ?

Quelle liberté restera-t-il aussi ?

Qualité des données

De mauvaises données au départ et une erreur peut se propager et tout le système sera faux, mais la décision sera prise... Attention danger !

Transhumanisme

Le transhumanisme¹⁵ est un mouvement culturel et intellectuel international prônant l'usage des sciences et des techniques afin d'améliorer les caractéristiques physiques et mentales des êtres humains.

Le transhumanisme, c'est améliorer l'homme, voire de préparer sa transformation grâce aux nouvelles technologies dites convergentes (nanotechnologies, biotechnologies, informatiques et cognitives).

Humanity+ a pour objectif :

- La promotion de l'amélioration de la condition humaine à travers des technologies d'amélioration de la vie, ayant pour but l'élimination du vieillissement et l'augmentation des capacités intellectuelles, physiques ou psychologiques.
- L'étude des bénéfices, dangers et de l'éthique du développement et de la mise en œuvre de ces technologies.

¹⁵ <https://iatranshumanisme.com/a-propos/transhumanisme/la-declaration-transhumaniste/quest-ce-que-le-transhumanisme-version-3-2/>

La liste ci-après est la traduction de la liste des sujets de base du transhumanisme.

Anti-âge	Amélioration de l'humain	Robotique
Intelligence Artificielle	Interface Homme-machine	Reprogénétique
Bionique	Technologie de l'information	Réalité simulée
Biotechnologies	Extension de la vie	Colonisation de l'espace
Interface cerveau-ordinateur	Chargement de l'esprit	Animation suspendue
Sciences cognitives	Transfert de l'esprit	« Wetware hacking »
Cryonique	Nanotechnologies	
« Cyberware »	Nanotechnologie moléculaire	
« Cyborg »	« Buckyballs »	
« Design » de bébé	Assemblage moléculaire	
Technologies émergentes	Piratage de neurone	
Ingénierie génétique	Développement personnel	
Clonage humain	Prothèses	
Clonage thérapeutique	Rajeunissement	

Transhumanist Technologies

Main article: List of basic transhumanism topics

Technologies which have become the focus of transhumanism:

- Anti-aging
- Artificial intelligence
- Bionics
- Biotechnology
- Brain-computer interface
- Cognitive science
- Cryonics
- Cyberware
- Cyborg
- Designer baby
- Emerging technologies
- Genetic engineering
- Cloning
 - Human cloning
 - Therapeutic cloning
- Human enhancement
- Human-machine interface
- Information technology
- Life extension
- Mind uploading
- Mind transfer
- Nanotechnology
 - Molecular nanotechnology
 - Buckyballs
 - Molecular assemblers
- Neurohacking
- Personal development
- Prosthetics
- Rejuvenation
- Robotics
- Reprogenetics
- Simulated reality
- Space colonization
- Suspended animation
- Wetware hacking

L'homme augmenté

Les sciences travaillent pour apporter du bien-être, du confort, du fonctionnel aux hommes ayant des défaillances (non-voyant, malentendant, membre coupé, etc.). Mais, si on changeait le corps d'une personne accidentée et que l'on ne gardait que la tête, serait-elle encore la même personne ?

Est-ce dans le cerveau que se trouve « la personne, son âme, sa conscience » ?

Ou, comme le dit Joël de Rosnay, « la vie est un courant léger » dans son livre *Je cherche à comprendre... les codes cachés de la nature*.

Arrivera-t-on à créer des machines avec une conscience ? C'est déjà fait...!

Quelques organismes

CNIL

La CNIL c'est protéger les données personnelles, accompagner l'innovation, préserver les libertés individuelles.

La CNIL organise un débat public décentralisé sur les enjeux de 2017 : algorithmes, Intelligence Artificielle et éthique jusqu'à l'automne. De nombreux forums et partenaires y participent. C'est ouvert à tous.

Comité Consultatif National d'Éthique – CCNE

Le Comité Consultatif National d'Éthique a pour mission de « donner des avis sur les problèmes éthiques et les questions de société soulevés par les progrès de la connaissance dans les domaines de la biologie, de la médecine et de la santé ».

Je ne suis pas certaine qu'il s'intéresse au numérique et à l'intelligence artificielle.

Future of Life Institute

Des organismes existent dans le monde pour prévenir les risques existentiels contre l'humanité.

Le « **Future of Life Institute**¹⁶ » est l'institut pour l'avenir de la vie. Y a-t-il donc danger ?

En janvier 2015, Bill Gates, Stephen Hawkins et Elon Musk (entre autres), dans cet article, nous disent que nous devrions avoir peur :

http://lexpansion.lexpress.fr/high-tech/intelligence-artificielle-attention-danger-meme-bill-gates-a-peur_1647411.html

Institut Convergence i2-Drive

I2-DRIVE signifie en anglais : Interdisciplinary Institute for Data Research: Intelligence, Values and Ethics, traduit en français : Institut Convergence pluridisciplinaire pour la recherche sur les données.

Il existe plusieurs instituts Convergences, financés par l'ANR et les investissements avenir.

¹⁶ https://fr.wikipedia.org/wiki/Future_of_Life_Institute

L'institut Convergence de l'université Paris-Saclay est coordonné scientifiquement par l'Inria. Ce projet collectif et interdisciplinaire se structure autour des défis scientifiques liés aux sciences des données et à leurs enjeux socio-économiques.

Dans son interview à L'Usine Nouvelle¹⁷, **Nozha Boujemaa**, directrice de recherche à l'Inria et conseillère du président de l'Inria sur le big data nous indique que l'intelligence artificielle a fait un bond important ces dernières années, que la qualité des données est très importante, que la transparence des algorithmes et les règles d'éthique sont nécessaires.

La plateforme **TransAlgo** doit évaluer la responsabilité et la transparence des systèmes algorithmiques.

Le principe est de développer et de diffuser des outils logiciels capables d'évaluer les algorithmes au regard de leur respect de la loi et des règles éthiques.

GAFIM

Google, Amazon, Facebook, IBM et Microsoft ont créé un partenariat en septembre 2016 qui œuvrera à améliorer la compréhension du public sur les technologies d'intelligence artificielle et à formuler les meilleures pratiques sur les défis et les opportunités de cette discipline.

https://www.lesechos.fr/28/09/2016/lesechos.fr/0211337588119_pour-une-intelligence-artificielle-responsable-et-sans-danger-pour-l-humanite.htm

https://www.lesechos.fr/28/09/2016/lesechos.fr/0211337588119_pour-une-intelligence-artificielle-responsable-et-sans-danger-pour-l-humanite.htm

Université de Stanford - AI100

Éric Horvitz¹⁸, Directeur de Recherche chez Microsoft, diplômé de médecine à l'université de Stanford, a lancé et financé (avec sa femme), fin 2014, une étude qui va durer **pendant un siècle**. L'objectif est d'étudier et anticiper les impacts de ces technologies sur tous les aspects de la société et de l'économie au travers des époques.

Le premier rapport de **AI100**¹⁹ est paru en septembre 2016. Une nouvelle étude sortira tous les 5 ans.

CONCLUSION

Les machines ont-elles une conscience ?

Est-ce qu'un jour les machines, leurs programmes et leurs données auront une conscience ? Est bien présomptueux celui qui l'affirmerait aujourd'hui...

J'ai trouvé la pyramide cognitive (voir le schéma ci-après) dans un tweet de Bernard Georges, chargé de prospective Stratégique à la Société Générale. Il semblerait que la conscience est l'étape suivante !

¹⁷ <http://www.usinenouvelle.com/editorial/intelligence-artificielle-la-non-veracite-des-donnees-est-un-risque-mortel-pour-nozha-boujemaa-directrice-de-recherche-a-l-inria.N543379>

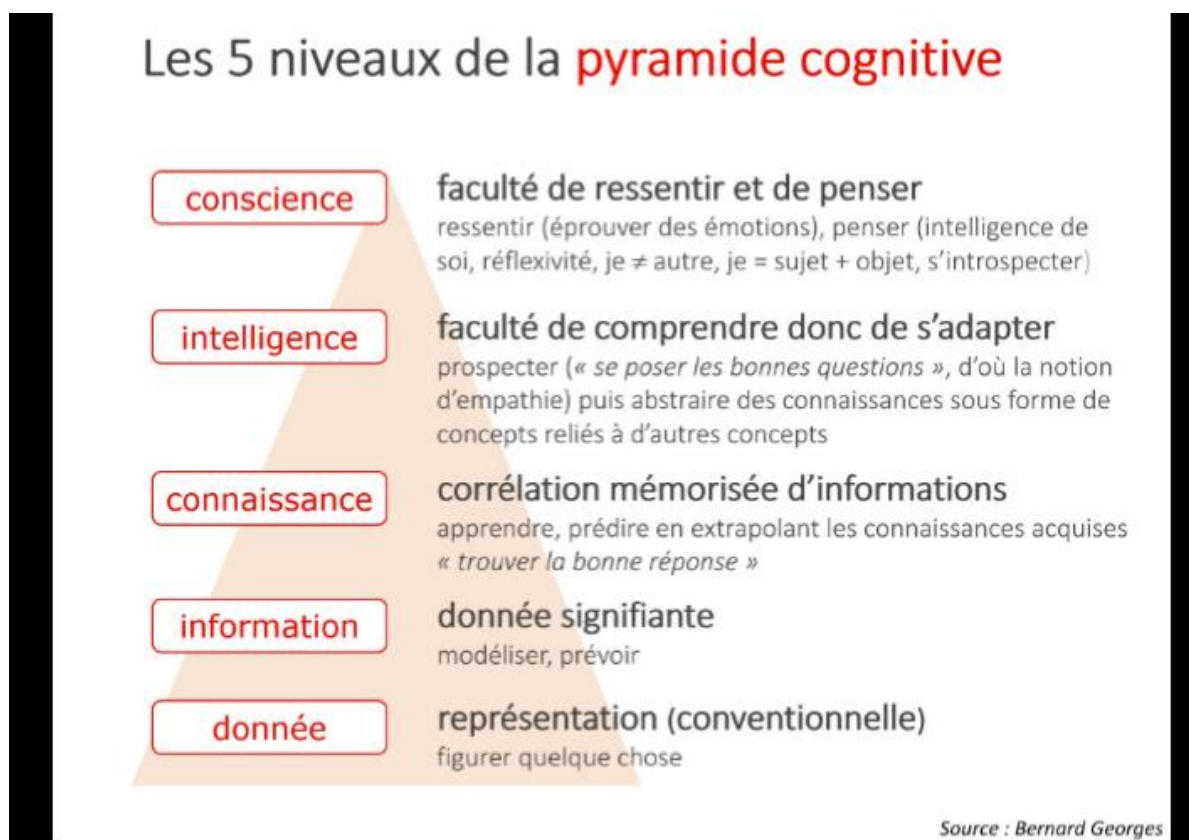
¹⁸ <https://qz.com/973005/microsofts-new-head-of-research-has-spent-his-career-building-powerful-ai-and-making-sure-its-safe/>
¹⁹ Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, Et Astro Teller. "Intelligence artificielle et vie en 2030." Étude de cent ans sur l'intelligence artificielle: rapport du panel d'étude 2015-2016, Stanford University, Stanford, Californie, septembre 2016. Doc: <http://ai100.stanford.edu/2016-report> . Accès: le 6 septembre 2016.



Cela interpelle et nous fait réfléchir. Est-ce vers cela que tendent les recherches d'aujourd'hui des grands acteurs du numérique ?

Soyons vigilants et ne laissons pas faire n'importe quoi avec les nouvelles technologies !

Une réflexion pluridisciplinaire sur ces sujets s'impose.



Bernard GEORGES @BdGEORGES · 24 juil. 2016

VIE AUGMENTÉE et EMPATHIE @derosnayjoel @changerdere @edgarmorinparis @dr_l_alexandre
youtu.be/vzr4e_WUV40 pic.twitter.com/M7sM3J3FGC

11 302 370

Une deuxième conclusion porteuse d'un peu d'espoir

PAI - « **Partnership on Artificial Intelligence to benefit people and society** » est une organisation mondiale qui recrute son directeur général. Le partenariat a commencé en février 2017 avec Amazon, Apple, Deepmind/google, IBM, Microsoft et Facebook. C'est un consortium désireux de définir les normes de l'industrie pour la transparence, la responsabilité et la sécurité pour les produits de l'IA.

Le partenariat est fondé sur les éléments suivants (traduits en français par Google - le texte original est au paragraphe suivant) :

En français

« Nous chercherons à faire en sorte que les technologies de l'IA bénéficient et autorisent autant de personnes que possible.

2. Nous éduquerons et écouterons le public et nous engagerons activement les parties prenantes à demander leurs commentaires sur notre attention, à les informer de notre travail et à répondre à leurs questions.
3. Nous nous sommes engagés à ouvrir des recherches et dialoguer sur les implications éthiques, sociales, économiques et juridiques de l'IA.
4. Nous croyons que les efforts de recherche et de développement de l'AI doivent être activement engagés et responsables envers un large éventail de parties prenantes.
5. Nous nous engageons et avons une représentation des intervenants dans le milieu des affaires pour nous assurer que les préoccupations et les opportunités spécifiques au domaine sont comprises et abordées.
6. Nous travaillerons à maximiser les avantages et à relever les défis potentiels des technologies de l'AI, en :
 - A. Travailler pour protéger la vie privée et la sécurité des individus.
 - B. S'efforcer de comprendre et de respecter les intérêts de toutes les parties susceptibles d'être influencés par les avancées d'AI.
 - C. S'efforçant de veiller à ce que les communautés de recherche et d'ingénierie de l'AI restent socialement responsables, sensibles et engagées directement avec les influences potentielles des technologies de l'IA sur la société dans son ensemble.
 - D. Veiller à ce que la recherche et la technologie de l'IA soient robustes, fiables, fiables et fonctionnent dans des limites sécurisées.
 - E. S'opposer au développement et à l'utilisation des technologies de l'IA qui violeraient les conventions internationales ou les droits de l'homme et la promotion de garanties et de technologies qui ne nuiraient pas.
7. Nous croyons qu'il est important que le fonctionnement des systèmes d'AI soit compréhensible et interprétable par les gens, afin d'expliquer la technologie.
8. Nous nous efforçons de créer une culture de coopération, de confiance et d'ouverture parmi les scientifiques et les ingénieurs d'AI pour nous aider à atteindre ces objectifs. »

Texte original en anglais

- « We will seek to ensure that AI technologies benefit and empower as many people as possible.
2. We will educate and listen to the public and actively engage stakeholders to seek their feedback on our focus, inform them of our work, and address their questions.
 3. We are committed to open research and dialog on the ethical, social, economic, and legal implications of AI.
 4. We believe that AI research and development efforts need to be actively engaged with and accountable to a broad range of stakeholders.
 5. We will engage with and have representation from stakeholders in the business community to help ensure that domain-specific concerns and opportunities are understood and addressed.
 6. We will work to maximize the benefits and address the potential challenges of AI technologies, by:
 - a. Working to protect the privacy and security of individuals.
 - b. Striving to understand and respect the interests of all parties that may be impacted by AI advances.



c. Working to ensure that AI research and engineering communities remain socially responsible, sensitive, and engaged directly with the potential influences of AI technologies on wider society.

d. Ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.

e. Opposing development and use of AI technologies that would violate international conventions or human rights, and promoting safeguards and technologies that do no harm.

7. We believe that it is important for the operation of AI systems to be understandable and interpretable by people, for purposes of explaining the technology.

8. We strive to create a culture of cooperation, trust, and openness among AI scientists and engineers to help us all better achieve these goals. »

Il faut réfléchir aux conséquences et à la façon de faire une IA responsable.

Qui est responsable quand un algorithme fournit des informations fausses ?

RÉFÉRENCES

Certaines de ces références sont en anglais.

<http://www.ladn.eu/entreprises-innovantes/parole-expert/seo-rankbrain-google-machine-learning-en-search/>

[https://www.lip6.fr/actualite/information-fiche.php?RECORD_KEY\(informations\)=id&id\(informations\)=18](https://www.lip6.fr/actualite/information-fiche.php?RECORD_KEY(informations)=id&id(informations)=18)

https://fr.wikipedia.org/wiki/Histoire_de_l%27intelligence_artificielle

http://www.lemonde.fr/pixels/article/2015/01/28/alan-turing-l-interminable-rehabilitation-d-un-genie_4564537_4408996.html

<https://fr.wikipedia.org/wiki/Prolog>

<https://fr.wikipedia.org/wiki/Lisp>

<http://lhomme.revues.org/18>

https://fr.wikipedia.org/wiki/%C3%89thique_de_l%27intelligence_artificielle

<http://cogprints.org/7284/>

<http://reinformation.tv/ethique-intelligence-artificielle-ia-jallais-65922-2/>

<http://www.4erevolution.com/the-future-of-life-institute/>

http://www.lemonde.fr/pixels/article/2015/07/27/intelligence-artificielle-hawking-musk-et-chomsky-reclament-l-interdiction-des-armes-autonomes_4701102_4408996.html

<https://code.google.com/archive/p/word2vec/>

https://fr.wikipedia.org/wiki/Apprentissage_profond

http://www.ionis-group.com/actualites/2017/05/grand_entretien_joel_de_rosnay_ionismag-35_ete_2017/#

http://gpp.oiq.qc.ca/qu_est-ce_que_l_ethique.htm

<http://internetactu.blog.lemonde.fr/2017/05/21/le-mythe-de-lia-surhumaine/>

<https://www.technologyreview.com/s/608108/forget-alphago-deepminds-has-a-more-interesting-step-towards-general-ai/>