



**HAL**  
open science

## Regards critiques sur l'Intelligence Artificielle, les intérêts politiques des empires numériques

Bernard Fallery

► **To cite this version:**

Bernard Fallery. Regards critiques sur l'Intelligence Artificielle, les intérêts politiques des empires numériques. 25<sup>ème</sup> Congrès de l'AIM, Jun 2020, Marrakech, Maroc. hal-03126059

**HAL Id: hal-03126059**

**<https://hal.science/hal-03126059>**

Submitted on 30 Jan 2021

**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.

# Regards critiques sur l'Intelligence Artificielle, les intérêts politiques des empires numériques

**Bernard FALLERY**

Université de Montpellier, MRM

Communication au 25ème congrès de l'AIM, Marrakech 2020

## Résumé

Depuis la peur d'une abolition du travail jusqu'aux prophéties sur l'abolition de la mort, un discours confus se développe autour de l'IA ; un discours qui mélange non seulement l'intelligence *Informatique* et l'intelligence *humaine*, mais aussi une intelligence *Fiction*. Pour éclaircir ces confusions une étude documentaire fait d'abord un point sur ces trois questions : quels sont les processus de l'IA-*Informatique* ? Quels sont les processus de l'intelligence humaine ? Quels sont les grands mythes de cette IA-*Fiction* ? La question posée devient alors : à qui profitent ces confusions qui apparaissent comme intentionnellement entretenues ?

La thèse défendue est que ce discours, entretenu par des malentendus sur l'IA, sert les intérêts des grands empires numériques, car il permet de distraire l'attention sur les vrais dangers de leur nouvelle puissance. Vrais dangers, car derrière ces confusions entre les réalités et les fictions semble se cacher deux visions politiques très opposées et qui, elles, ne sont pas du tout ambiguës : d'un côté la vision des libertariens nord-américains pour un absolu contrôle de soi-même en face aux institutions d'un état démocratique et de l'autre côté la vision du capitalisme d'état chinois pour un contrôle absolu des citoyens en empêchant l'apparition des structures d'un état démocratique.

**Mots clés** : intelligence artificielle, discours, transhumanisme, gouvernementalité

## Abstract

From the fear of an abolition of labor to the dream of an abolition of death, a whole confused discourse develops around AI; a discourse that mixes not only computer intelligence and human intelligence, but more and more an intelligence fiction. To clarify these confusions, a documentary study first gives an update on these three questions: What are the processes of the AI-Informatics? What are the processes of human intelligence? What are the big myths of this AI-Fiction? The question then becomes: who benefits from these confusions that appear as intentionally maintained? The argument is that this discourse, sustained by AI misunderstandings, serves the interests of the great digital empires, as it distracts attention from the real dangers of their new power. Real dangers, because behind these confusions between the realities and the fictions are hidden two very opposite political visions and which, they, are not at all ambiguous: on the one hand the vision of the North American libertarians for an absolute control over oneself - even by gradually attacking all the institutions of a democratic state - and on the other hand the vision of Chinese state capitalism for an absolute control of the citizens - by blocking any appearance of the structures of a democratic state.

**Keywords** : artificial intelligence, human intelligence, transhumanism, governmentality

## De l'intérêt scientifique des néologismes ambigus

Lorsque J. McCarthy et M. Minsky utilisent en 1956 le néologisme ambigu d'Intelligence Artificielle lors d'un séminaire à l'université de Dartmouth ([Prade 2016](#)), ils comprennent rapidement que cela peut attirer l'attention du public et faciliter leurs recherches originales sur la modélisation des fonctions cognitives (les véritables sujets, sur la logique formelle et sur la

représentation des connaissances, étaient certes moins alléchants). Après trop peu de réalisations et vingt ans plus tard ce sera au néologisme tout aussi ambigu de Système Expert de jouer le même rôle, pour financer cette fois les recherches sur la résolution de problèmes et sur le diagnostic : ici aussi, les intitulés des véritables concepts (ceux de logique propositionnelle, de base de connaissances et de moteur d'inférence, bien décrits par exemple sur le site [Will-Bridge](#)) n'auraient sans doute pas autant attiré d'investissements.

### **Du danger politique des néologismes ambigus**

Aujourd'hui c'est de nouveau l'étiquette IA qui est revenue en force. Ses succès actuels permettent bien sûr de financer de très gros budgets, publics et privés (Rapport du pôle [Bipame](#) 2019), mais ce n'est pas ici le plus intéressant. Car l'étiquette IA est devenue encore plus ambiguë qu'en 1956, puisqu'elle mélange maintenant non seulement l'IA *Informatique* et l'intelligence *Humaine*, mais aussi une IA *Fiction*. Et comme on peut le constater dans les médias, ces confusions sont savamment entretenues par les nouveaux empires numériques que sont devenus les GAFAMI nord-américains et les BHATX chinois ; il est donc temps de se demander à qui peuvent profiter ces confusions.

Pour le comprendre il faut accepter de parler d'intelligences au pluriel et quatre questions sont alors ici développées :

- comprendre où en est aujourd'hui l'intelligence artificielle *Informatique* est nécessaire pour permettre de discuter des enjeux à venir, puisque les opportunités et les risques sont effectivement devenus impressionnants ;
- mais revenir (encore une fois) sur les processus de l'intelligence *Humaine* est nécessaire, ne serait-ce que pour...
- garder une vision claire sur cette intelligence artificielle *Fiction*, dont l'analyse est elle aussi passionnante, quand on considère bien qu'elle n'est pas de l'ordre des faits mais des grands mythes de l'humanité ;
- pour au final mieux comprendre à qui profitent aujourd'hui toutes ces ambiguïtés et même ces confusions entretenues entre l'informatique, la personne humaine et la fiction.

D'un point de vue méthodologique ce travail repose sur une étude documentaire du « discours sur l'IA ». La lecture de documents s'est poursuivie jusqu'à obtenir saturation dans les thèmes abordés. Pour l'analyse elle-même, 39 publications académiques et 55 publications d'actualité ont finalement été retenues. Le codage a été réalisé à la main sous la forme de surlignages et d'annotations par mots-clés, puis les mots-clés ont pu être regroupés en quatre grands thèmes, chacun autour de sous-thèmes qui organisent cette présentation : intelligence informatique (IA numérique, IA symbolique), intelligence humaine (créativité, esprit critique), intelligence fiction (point de singularité, transhumanisme) et peurs/dangers (Silicon Valley et capitalisme chinois). Pour ne pas surcharger la présentation, les verbatims ne sont pas recopiés dans le texte mais la liste en bibliographie présente tous les liens actifs.

## **1. Les processus de l'IA-informatique : reconnaissance artificielle et logique artificielle**

Il n'est pas facile de mettre un peu d'ordre dans la profusion actuelle des contributions associées à l'étiquette IA-Informatique, depuis par exemple le nouvel Assistant de Google qui peut téléphoner à notre place pour prendre rendez-vous avec un coiffeur (Vidéo [Youtube](#) 2018) jusqu'à l'hypothèse d'une dérive malveillante lors de la mise en résonance involontaire de plusieurs systèmes de robots militaires ([Berthier et Kempf](#) 2017).

En informatique, l'IA a pour objet de décomposer des fonctions cognitives pour simuler des comportements humains dans ses différentes activités : perception-acquisition, mémoire-

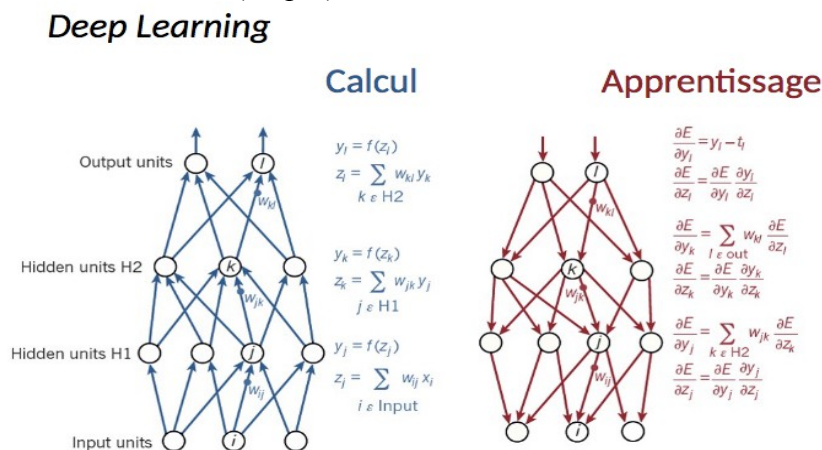
apprentissage, raisonnement-pensée, expression-communication et exécution-décision. Il s'agit bien de simulation, de reproduire des résultats et non des processus humains. Mais derrière ce programme général il faut distinguer l'approche de l'IA *numérique* (qui s'appuie sur les progrès de la statistique et des bases de données massives) et l'approche de l'IA *symbolique* (qui s'appuie sur les progrès de la logique formelle et de la représentation de connaissances). On peut alors mieux comprendre les risques de « disruptions » sociales, aujourd'hui liés à la conjonction entre apprentissage automatique et données massives.

## 1.1 L'IA numérique : la reconnaissance artificielle

Depuis une quinzaine d'années c'est l'approche de l'IA *numérique* (ou connexionniste) qui domine la discipline ([MIT Technology Review](#) 2019), par une conjonction entre trois types d'évolutions :

- des évolutions proprement informatiques : sur la gestion des données (bases de données massives et distribuées, *NoSQL*, *Hadoop*, *Cloud computing*...) et sur la puissance de calcul (calcul parallèle sur des grappes de serveurs *HPC*, coût et miniaturisation des nouveaux capteurs et des processeurs graphiques) ;
- des évolutions en statistiques sur la classification automatique, qui ont permis la fouille de données : processus de décision markoviens, algorithmes des k plus proches voisins, régressions, forêts d'arbres de décision, partitionnement de données similaires, calculs de singularités...
- et aujourd'hui des progrès dans l'utilisation des réseaux de neurones formels : le Perceptron, première conceptualisation d'un neurone formel apprenant, avait été proposé par F. Rosenblatt en 1957 (Cours [Youtube](#) 2010) mais c'est en 2019 que les travaux sur les architectures de réseaux de neurones ont valu le prix Turing à Y. LeCun , Y. Bengio et G. Hinton ([LeCun](#) 2016)

Un réseau de neurones formels est en fait une succession de matrices dont les coefficients sont les poids « synaptiques » des neurones. L'application d'un vecteur en entrée d'une matrice (stimulus) entraîne une modification de la valeur de chaque coefficient suivant la moyenne pondérée des entrées et une fonction de seuil ; la modification se transmet aux coefficients connectés de la matrice suivante, jusqu'à obtenir un vecteur en sortie (output).



Lecun et al., Deep Learning, *Nature Review* 2015

Il s'agit alors d'entraîner le réseau constitué, qui peut *apprendre* en fonction des erreurs en sortie, par une « retro-propagation » pour les corriger : c'est le principe du *Machine Learning*. Le concepteur du réseau doit notamment choisir le meilleur pas de correction des poids synaptiques (algorithme de minimisation du gradient d'erreur, de Rumelhart-

Hinton-Williams, [Ezratty](#) 2018, page 56) pour qu'ensuite ces poids synaptiques *se règlent d'eux-mêmes* de manière optimale et stable ([P. Paquet](#) 1997 donne un exemple en finance et le livre blanc de [Smile](#) 2018 présente différents kits de développement actuels en *Machine Learning*).

Un tel réseau hiérarchique à couches multiples de matrices peut alors *raisonner* grâce à un *calcul de distance* entre des vecteurs : [Y. LeCun](#) (2016) parle de « vectorisation du monde » (pour la conduite autonome le vecteur Moto est proche du vecteur Voiture et ils sont différents du vecteur Piéton), une approche très différente de la « modélisation du monde » envisagée par les ontologies formelles, comme on le verra plus loin.

L'entraînement d'un réseau se fait donc uniquement « par expérience », soit en apprentissage supervisé sur des données déjà labellisées, soit en apprentissage par renforcement suivant des valeurs de récompenses. La masse considérable de données nécessaires devenant très coûteuse en puissance de calcul, les recherches portent autant sur des processeurs spécialisés que sur de nouvelles architectures de matrices. Certaines architectures complexes amènent alors [Y. Bengio](#) (2019) à parler de « la révolution de l'apprentissage profond » (c'est la structure des réseaux qui est profonde) :

- les réseaux convolutifs ont par exemple la particularité d'utiliser une hiérarchie de matrices-filtres (pour ne traiter qu'une portion de l'information) qui génèrent des matrices-maps ou patterns ([LeCun](#) 2016) : ces réseaux sont alors capables distinguer des formes avec des *niveaux d'abstraction successifs*, pour analyser *pas à pas* une image ou un son ;

- les réseaux récurrents ont par exemple la particularité d'avoir une structure qui n'est pas uniquement linéaire, mais qui contient aussi des connexions sous forme d'arcs ou cycles de retro-action : en pouvant alors mémoriser et *réinjecter des entrées passées*, ces réseaux permettent l'analyse ou la construction de séries temporelles (les phonèmes pour la voix, les mots pour le langage, les notes pour une partition...).

- les réseaux génératifs adverses GAN ont par exemple la particularité de générer des formes et de les soumettre à un réseau discriminateur qui peut rejeter celles qui ne lui apparaissent pas réelles : le processus d'apprentissage s'arrête quand le Générateur est capable de créer des formes qui trompent le Discriminateur à tous les coups ([Wintics](#) 2018).

Une bonne source d'information sur ces applications actuelles est l'ebook en accès libre de [O. Ezratty](#) (2018, 522 pages) qui présente ces nouvelles architectures de réseaux de neurones mais aussi les implications pour les transports, le bâtiment, les télécoms, l'agro-alimentaire, la distribution, la finance, la santé...

Par la « simple » perception-analyse d'une masse de situations ponctuelles dans un domaine particulier, les réseaux de neurones formels infèrent notamment des fonctions de prédiction *par induction* très intégrées (estimations boursières, profils de clients, prédiction des fraudes, analyse financière, assistants virtuels Chatbots, systèmes de recommandation, tarification discriminante en temps réel, conduite autonome...), mais ils ne fournissent aucune règles ou justifications *pour l'interprétation*, puisque les résultats obtenus en sortie ne sont pas explicables : quand le réseau de matrices devient « profond » (*Deep learning* : jusqu'à 150 matrices dans le modèle Resnet, [Microsoft Research](#) 2015), les millions de modifications automatiques du potentiel de tous les neurones ne sont plus *traçables*.

Cette approche numérique ou connexionniste de l'IA connaît des succès spectaculaires dans les domaines où on peut s'appuyer sur un *entraînement* suffisamment rapide d'un réseau de neurones et sur une masse considérable d'exemples : vision et reconnaissance d'images, reconnaissance faciale, reconnaissance de la voix et des émotions, reconnaissance du langage naturel et traduction,

reconnaissance des perturbations génétiques... On voit ainsi apparaître la formule : IA = *Big data* + *Machine learning* (ce dernier représentant en effet 89% des 55.000 brevets déposés en IA en 2017, [OMPI](#)).

## 1.2 L'IA symbolique : la logique artificielle

C'est pourtant l'approche de l'IA *symbolique* (ou logique) qui avait fondé la discipline ([Ezratty 2018](#)), d'une part en considérant le raisonnement comme une manipulation de symboles s'appuyant sur la logique formelle (logique des propositions vraies ou fausses, logique des prédicats sur des variables quantifiables, logique floue avec des degrés de vérité...) et d'autre part en rendant possible une représentation du monde sous la forme d'ontologies formelles :

- il ne s'agit pas ici de reconnaître des formes et d'extraire des relations prédictives, il s'agit de reproduire un processus mental peu structuré en se basant sur des structures informatiques différentes des bases de données classiques : des ontologies formelles d'un domaine de connaissances, des réseaux et graphes conceptuels, des *frames*, des bases de règles, des systèmes multi-agents... Il s'agit donc de résolution de problèmes en exploitant des connaissances existantes dans des domaines déjà théorisés : démonstrateurs de théorèmes, moteurs de recherche sémantiques, parcours des arbres de décision, extraction de l'expertise... et un des intérêts majeurs pour l'aide à la décision est de pouvoir fournir ici une *explication* à la solution informatique proposée pour résoudre le problème : par exemple par la trace des règles utilisées par un moteur d'inférence dans un système expert ou par l'enchaînement des échanges entre entités dans un système multi-agents ([Ferber 1995](#)) ;
- d'une part cette approche symbolique ou logique de l'IA n'est certes pas adaptée à la fonction de perception-reconnaissance et d'autre part les connaissances restent difficiles à représenter et à conceptualiser : elles sont à la fois tacites et explicites, et elles nécessitent des codages et des mises à jour qui sont encore fastidieux. Mais de nombreux chercheurs plaident néanmoins aujourd'hui pour son renouveau et son association avec l'approche numérique, dans un contexte où le *Machine Learning* uniquement statistique (toujours plus de données, toujours plus de processeurs) semble commencer à montrer les limites d'une « intelligence superficielle » qui ne peut pas intégrer un minimum de sens commun sous la forme de savoirs a priori (règles, modèles prédictifs, inférences...), [InternetActu 2017](#), [L'UsineNouvelle 2018](#). On se souvient par exemple de l'embrassement général après la victoire au jeu de Go ([Medium 2017](#)) d'AlphaGo Zero de DeepMind contre Lee Sedol... mais la machine a bizarrement pris sa retraite ([Znet 2017](#)) et on a appris en 2019 que cette IA de [DeepMind](#), n'utilisant toujours que *l'apprentissage par renforcement statistique*, s'est révélée incapable de passer un contrôle de mathématiques de niveau lycée pour ne pas savoir mettre un problème sous forme d'équation ([Le Figaro 2019](#)).

## 1.3 L'écosystème de l'IA et les risques de disruption sociale

Autour du « pivot » que constitue en ce moment l'IA numérique avec le *Machine Learning* et les *Big data*, ce sont les risques de disruption qui guettent la « périphérie » sociale. Par disruption il faut considérer ici un saut dans les étapes d'un développement, un court-circuit sans délibération des modèles établis, une transgression. Par risques il faut considérer ici ceux qui sont liés aux nouvelles formes que prennent la rationalisation, l'automatisation et le « mode de gouvernement des individus » :

- le risque de la *rationalisation* est celui d'une disruption autour du vieux rêve managérial du taylorisme. Par la conjonction du *Machine learning* et des *Big data*, la rationalisation prend en effet aujourd'hui la forme d'une véritable quantification du monde, d'un impérialisme des données, d'une dictature des classements ([Gori et Del Volgo 2009](#)) et d'un océan de

corrélations ([Boyd et Crawford](#) 2011). Mais en constatant la difficulté croissante à rendre les algorithmes redevables des décisions prises (on a vu que la correction automatique du poids des neurones formels ne devient plus vraiment traçable) et en constatant d'autre part la gravité des externalités négatives quand une personne est concernée par les résultats (discriminations dans l'accord d'un prêt, dans la sélection de candidats, dans les jugements de libération conditionnelle, dans la pose d'un diagnostic médical... [Zuiderveen-Borgesius](#) 2018), le risque le plus pressant aujourd'hui deviendrait moins aujourd'hui l'intelligence artificielle que l'absence de contrôle d'une « bêtise » artificielle ([Benavent](#) 2016).

- le risque de *l'automatisation* est celui d'une disruption autour du vieux rêve économique du capitalisme de la maîtrise du facteur travail. Même si c'est plutôt l'ensemble du progrès technologique qui modifie notre vie (et ce depuis le plus basique des téléphones portables), l'IA fait apparaître à la fois de nouvelles opportunités (liées à la productivité du capital, de l'énergie et des matières premières, Rapport [Lemoine](#) 2014) mais aussi de nouveaux risques (liés notamment à la marginalisation du facteur travail dans de nombreux emplois de « cols blancs » quand le travail est mono-tâche, [Ezratty](#) 2016 page 421). Sans parler des caissières ou des chauffeurs, et selon une étude d'Accenture, plus de 50% du temps de travail des managers serait concerné ([Harvard Business Review](#) 2016). Quant au travail actuel d'un radiologue sur la reconnaissance des tumeurs, il aura sans doute disparu dans une dizaine d'années (actes du séminaire de la [FNMR](#) 2018). Les prévisions globales sur ces modifications de l'emploi sont très contradictoires, mais ce qui est aujourd'hui clair c'est déjà un transfert de travail vers les clients (caisses automatiques, centres d'appels, banques en ligne...), un transfert de travail vers de nouveaux précaires (qui sont sortis du salariat pour être « ubérisés ») et un transfert de travail vers tous les nouveaux « micro-travailleurs du clic » indispensables à l'IA numérique à l'exemple de la plateforme de micro-tâches Amazon Turc Mécanique (dans ce *digital labour*, on peut imposer d'analyser jusqu'à 2000 images par heure, [Casilli](#) 2018).

- enfin le risque *sociétal* est celui d'une disruption dans la « gouvernamentalité », décrite par M. Foucault comme associant un mode de gouvernement des individus et une rationalité ([Foucault](#) 1978). Une gouvernamentalité algorithmique ([Rouvray](#) 2012) devient maintenant le mode de gouvernement des personnes :

- qui n'opère plus par une réglementation des conduites, mais par la configuration anticipative des possibles et une incitation à l'identique dans des environnements personnalisés à l'avance (Filtres Google, Fils d'info Facebook...);
- où le pouvoir ne s'exerce non pas seulement par cette connaissance du probable, mais par une maîtrise sur le potentiel (Recommandations Netflix, Tendances Twitter, Notes TripAdvisor ...) sur les opportunités (les *nugdes*, coups de coude discrets qui orientent les comportements...) et sur les réflexes (alertes sur les objets connectés...);
- un mode de gouvernement qui, au delà même d'une soumission volontaire, relève presque d'une servitude bienheureuse : les données personnelles ne sont plus privées, mais elles appartiennent à de nouveaux seigneurs, en échange de services « gratuits » (pour lesquels Waze a remplacé le moulin ou le four à pain communal);
- un mode de gouvernement qui relève à la fois du « techno-mimétisme » (où l'individu aligne inconsciemment son comportement afin de pouvoir utiliser efficacement l'Intelligence Informatique, [Marosan](#) 2019) et de la « sousveillance » (où l'individu n'est même plus surveillé mais plutôt sous-veillé par ses traces numériques, de façon discrète, immatérielle et omniprésente, [Sadin](#) 2009, [Quessada](#) 2010).

Et si on a envie de se faire vraiment peur, on peut lire ce témoignage dans [Mediapart](#) 2018 sur la gouvernamentalité « à la chinoise ».

## 2. Les processus de l'intelligence humaine : créativité et esprit critique

Les progrès en neurosciences éclairent de plus en plus le côté fascinant du fonctionnement du cerveau humain, cet organe de 80 milliards de neurones dotés chacun de 10.000 synapses et qui peut être considéré comme le résultat le plus sophistiqué issu de l'évolution puisqu'il est même capable de comprendre jusqu'à son propre fonctionnement ([Dubuc 2018](#), conférences à McGill). Mais si nous pouvons de mieux en mieux expliquer, par exemple, les liens neurobiologiques entre raison et émotion ([Damasio 2010](#)), il reste toujours aussi difficile de définir les différentes formes de l'intelligence humaine : la pensée n'est-elle qu'une forme de calcul ? Peut-on définir différents degrés de la conscience ? Le système nerveux de notre estomac est-il intelligent ? Quelle est la part de l'intelligence biologiquement héritée et celle de l'intelligence socialement acquise ? Peut-on tenter une formalisation mathématique de l'intelligence générale, basée sur la complexité et la concision ?... En se limitant ici à la question de savoir si les processus de l'intelligence Humaine devraient craindre la concurrence des puissants processus de l'intelligence Informatique, deux points apparaissent essentiels : la créativité et l'esprit critique.

### 2.1 La créativité, comme différence entre « savoir reconnaître » et « créer une connaissance »

Pour savoir si une machine fait montre d'intelligence, on connaît le succès de la proposition d'[Alan Turing](#) (1950) : un test à l'aveugle pour attribuer un dialogue à une machine ou à un humain, test donc uniquement basé sur la performance d'une imitation ressentie... et évitant ainsi justement d'avoir à définir l'intelligence. Même si ce « jeu de l'imitation » basé sur le langage s'est révélé être une incitation considérable pour le développement de l'IA, il a aussi fait oublier le fait qu'un même résultat intelligent et inventif (ce que peut faire une machine et de mieux en mieux : écrire une [partition musicale](#), dessiner un [tableau](#), composer un [roman](#), imiter une [voix](#), créer une page [Wikipedia](#), tenir une [conversation](#) en reconnaissant des [émotions](#), créer de [l'empathie](#) avec un chatbot...) n'implique pas des *processus identiques* pour l'obtenir :

- on a vu qu'en IA un système apprend et raisonne, soit en manipulant logiquement des symboles (dès 1963, c'était déjà la proposition de [A. Newell et H. Simon](#)), soit en manipulant numériquement des vecteurs et des matrices (c'est aujourd'hui la proposition de l'IA numérique). Un ordinateur peut donc être *inventif* dans la mesure où le résultat obtenu n'est non seulement pas prévisible mais peut même être entièrement généré, comme c'est le cas par exemple pour des images, des textes ou des architectures avec la nouvelle approche des réseaux de neurones GAN Generative Adversarial Network ([Wintics 2018](#)). Un robot peut donc être *autonome* dans la mesure où la régulation se fait autant par rapport à son état interne que par rapport à son environnement, on parle d'autorégulation ou même d'auto-organisation non prévue par le programme de départ ;
- mais un être humain a non seulement conscience de lui-même (comme certains animaux ou même certains robots « zombies » qui réussissent le « test du miroir », [Interstices 2016](#)), mais il *a aussi* conscience de penser, et *surtout* il le fait dans un contexte social, psychologique et temporel :
  - conscience de penser dans un contexte social, en intégrant non seulement ses propres expériences *mais aussi* des connaissances de tiers de confiance et des controverses dans l'espace public ;
  - conscience de penser dans un contexte psychologique et corporel, en faisant interagir non seulement plusieurs représentations qui sont connues *mais aussi* des sentiments, des souvenirs et des introspections qui associent le réel et l'imaginaire ;
  - conscience de penser un contexte temporel, selon diverses logiques et objectifs qui



peuvent être non seulement subis *mais aussi* qui peuvent être choisis, initiés ou adaptés selon le déroulement des circonstances.

Si un algorithme peut certes faire des découvertes étonnantes en exploitant des bases de données massives, il s'agit essentiellement de reconnaître des régularités ou des singularités. Un exemple : c'est le séquençage de l'ADN des échantillons prélevés lors de l'expédition Tara Océans qui a révélé plusieurs milliers d'espèces de plancton qui étaient inconnues et dont un tiers ne peuvent même pas être rattachées à un groupe répertorié ([Pour la Science](#) 2019).

La créativité quant à elle est en revanche toujours le fruit de l'interaction avec le contexte, « un processus de conception d'une solution jugée nouvelle, innovante et pertinente *en lien* au contexte précis de la situation-problème » ([Romero et al. 2017](#)). Mais le plus important est que ce contexte précis est bien plus qu'un « environnement », il est bien sûr socio-culturel et socio-psychologique, mais il est même socio-historique : la vérification des théories sur les ondes gravitationnelles ou sur le boson de Higgs s'est ainsi étalée sur un siècle ([La Recherche](#) 2016). Quant à l'étage supérieur, c'est-à-dire celui d'un changement de paradigme, d'un bouleversement de la perspective sur le monde par l'association de nouveaux concepts, on comprend alors que cela n'a aucun sens pour une intelligence artificielle, même si elle devient un jour à la fois numérique et symbolique. La cognition humaine ne se résume pas à un simple empirisme perceptif en additionnant des images, et pour reprendre une jolie formule de [D. Cardon](#) (2015) « il est encore temps de dire aux algorithmes que nous ne sommes pas la somme imprécise et incomplète de nos comportements ».

### 2.3 L'esprit critique, comme antidote à la « connerie humaine »

En 1983 dans son ouvrage *Où en est la psychologie de l'enfant*, R. Zazzo a pu se permettre de reformuler la question de l'intelligence sous la forme « *Mais qu'est-ce que la connerie, madame?* » (in [Beaumat](#) 2008), car pour lui « le contraire de la connerie, ce n'est pas la logique » (Zazzo 1983, page 47). En 2018, [C. Hadji](#) (2018) peut se permettre d'expliquer que « la tension qui oppose l'insuffisance de maîtrise à l'efficacité technique, et pour laquelle l'IA serait le pôle supérieur, ne se superpose pas avec une autre tension, celle qui oppose connerie et intelligence critique ».

« *Entre nous soit dit, bonnes gens/Pour reconnaître/Que l'on n'est pas intelligent/Il faudrait l'être* » (G. Brassens, dans *Ceux qui ne pensent pas comme nous*) : on voit que pour *prendre conscience* de ses insuffisances il faudrait déjà être capable d'esprit critique, c'est à dire savoir s'interroger sur la valeur et les conséquences de ses actes.

L'intelligence humaine est plurielle et suivant l'usage de nos capacités cognitives « on peut à la fois être con et intelligent » (Zazzo, 1983, page 48). C. Hadji (2018) donne comme exemple de « se mettre à quinze pour tabasser un lycéen sans défense » : la conscience critique « cette autre dimension de l'intelligence, serait en quelque sorte un antidote pour la connerie » (et alors les machines, ne pouvant exercer qu'une intelligence logico-mathématique, seraient définitivement à l'abri d'une connerie artificielle).

Le véritable danger pour l'intelligence Humaine ne serait alors pas aujourd'hui une concurrence déloyale de l'intelligence Informatique, mais au contraire celui d'un abandon de notre esprit critique et de tout libre arbitre ([Scaruffi](#) 2016) : non seulement l'homme risque la bêtise en se déchargeant de plus en plus de nombreuses fonctions intellectuelles comme la mémoire, le calcul, la lecture... ([Desmurget](#) 2019), mais il risque surtout la connerie, l'absence d'esprit critique, en imitant l'intelligence numérique ou logique de l'IA. Car finalement, ne devrait-on pas inverser la question du fameux jeu de l'imitation de Turing et se poser cette question nouvelle : aujourd'hui, entre un humain et une machine, Qui imite Qui ? ([Marosan](#), 2019). Ne sommes-nous pas en train d'intérioriser, sous la pression d'une optimisation ressentie comme nécessaire, une bonne partie des processus de l'intelligence artificielle : adapter son langage pour se faire mieux comprendre par son assistant vocal, adapter ses choix en fonction des recommandations Netflix ou Amazon, adapter ses

mouvements pour améliorer un score de santé sur sa montre connectée, adapter sa lecture des actualités en fonction de son fil d'info Facebook, adapter ses conversations au rythme du multi-tâche imposé par ce téléphone mobile omniprésent... les êtres humains, eux, savent se faire robots.

Au final, et de la même manière que personne n'a eu l'idée d'appeler un avion un « oiseau artificiel », on dira que l'intelligence artificielle fonctionne suivant d'autres processus et qu'elle n'a besoin ni de créativité ni d'esprit critique pour être intelligente à sa manière. A l'avenir, l'IA sera sûrement composée de nombreuses IA différentes, dans des domaines spécifiques et avec des capacités spécifiques en perception, apprentissage, raisonnement, communication, exécution. Ces capacités sont et seront souvent supérieures aux capacités humaines dans des domaines particuliers, mais cela ne mène pas à une intelligence reposant sur la créativité et l'esprit critique. Pour reprendre notre parallèle, personne ne se demande quand l'intelligence d'un avion aura dépassé l'intelligence d'un oiseau.

D'un point de vue éthique, le plus important ne serait donc sans doute pas de se préoccuper d'Une Super-IA potentiellement plus avancée que nous (même si cela peut être bien excitant), mais plutôt des défis que poseront la complémentarité de différentes IA avec l'humain (même si cela est plus terre à terre, comme le montrent par exemple [les vifs débats](#) sur une éventuelle personnalité juridique des robots). Lancée en 2018 par l'Université de Montréal, la [Déclaration de Montréal](#) définit les principes de l'IA responsable, mais la régulation dans les entreprises ou par les états est encore au stade des promesses (H. Guillaud sur [Internetactu](#) 2019).

### **3. Les grands mythes sur la vie et sur la mort : l'IA-fiction**

Parmi les grands mythes qui ont toujours fasciné et façonné l'humanité, il y a bien sûr les mythes sur le pouvoir de créer la vie et sur le pouvoir de repousser la mort. Ce n'est pas parce que ces mythes ne sont pas « l'objet » de la science qu'il faut les évacuer d'une réflexion scientifique. Et dans cette réflexion sur l'intelligence artificielle, revenir sur ces mythes permet alors de comprendre les relations entre les peurs entretenues sur l'IA-Fiction et les dangers politiques qui s'appuient sur l'IA-Informatique.

#### **3.1 Le mythe de la singularité, nouvelle version de Prométhée, de Pygmalion, du Golem...**

Prométhée a volé le feu sacré de l'Olympe pour en faire don aux humains qui n'avaient rien, ni la force des lions ni les ailes des oiseaux. Feu divin, feu connaissance, technologie primordiale mais dangereuse. Prométhée le titan sera puni par Zeus, son foie sera dévoré chaque jour par l'Aigle du Caucase, car son énorme ambition pourrait amener les humains à se surpasser. Morale de la fable : le rêve des humains qui voudraient que la technologie les fasse entrer dans le monde des dieux ne date pas d'hier.

Pygmalion est tellement fasciné par la perfection de cette statue qu'il a longtemps modelée, qu'il obtient d'Aphrodite qu'elle lui donne la vie. Il en tombe amoureux et épouse cette femme qu'il a créée de ses mains, devenue Galatée. Morale de la fable : le rêve des humains fascinés par la frontière entre eux et les objets qu'ils créent ne date pas d'hier.

Le Golem est apparu au XVI<sup>e</sup> siècle dans une communauté juive de Prague qui vivait dans la peur des calomnies. Pour se protéger et sur recommandation divine, le sage Yehoudah Loew sculpta alors un homme dans la glaise de la rivière avant de lui insuffler la vie selon un rituel secret. Mais sa création finit par se retourner contre ses maîtres et le sage dut lui ôter la vie. Morale de la fable : la fascination des humains face au pouvoir à la fois bénéfique et maléfique d'objets créés à leur image ne date pas d'hier.

Il y a fort à parier que Dmitry Itskov, entrepreneur russe qui a lancé son « Initiative 2045 » a du être fasciné par ces mythes ancestraux sur la maîtrise de la vie. Pourquoi 2045 ? Parce que les calculs montrent que c'est à cette date que le « point de singularité technologique » sera atteint. Sur le site <http://2045.com/> on peut donc voir toutes les étapes qui vont nous amener en 2045 à l'apparition de l'avatar de type C, celui dont l'intelligence commencera à dépasser l'intelligence humaine puisque nous pourrions y télécharger notre propre esprit. Le site affiche plus de 47.000 abonnés et non des moindres puisqu'il compte en autres le Dalai-lama et K. Kurzweil, un des nombreux directeurs du développement chez Google.

La « démonstration » du point de singularité est en effet le cheval de bataille de K. Kurzweil ([Conférence TED](#), la [Singularity University](#)). En mathématique un point de singularité est un point critique pour une fonction (par exemple,  $1/x$  quand  $x$  tend vers zéro), mais ici il ne s'agit que du point d'intersection de deux courbes: celle du développement de l'intelligence humaine, qui est croissante depuis la préhistoire mais pratiquement linéaire, et celle du développement de l'intelligence artificielle qui est exponentielle depuis 1950. Les deux courbes vont donc se croiser en 2045, l'argument étant que les capacités de l'IA vont suivre une loi identique à la fameuse « loi de Moore » sur le doublement tous les deux ans de la puissance des microprocesseurs (voir Cerebras le plus grand processeur d'IA, [Ozeratty 2019](#)). On passera ainsi de l'IA faible d'aujourd'hui (*Artificial Weak Intelligence*) à l'IA Forte de 2045 (*Artificial General Intelligence*), avant de voir s'épanouir bientôt la Super-IA (*Artificial Super Intelligence*).

La discussion de cette thèse, simple et spectaculaire, agite les médias en quête d'audience puisqu'elle donne des frissons à ceux qui aiment se faire peur avec la technologie : de l'IA Allégresse... jusqu'à IA Apocalypse. Même une église de l'IA est née en 2017, [WayOfTheFuture](#). Mais l'hystérie prophétique de la Super-IA agite aussi une partie des experts : quelques grands scientifiques dont l'informatique n'est pas le domaine, tel fut le cas de Stephen Hawking, mais surtout des ingénieurs travaillant dans les GAFAMI qui se donnent le vertige d'appartenir aux géants du numérique qui pourraient bouleverser l'humanité.

L'analyse de J-G Ganascia ([Ganascia 2017](#)) sur ce mythe de la singularité permet de garder la tête un peu plus froide ([Conférence](#) Montpellier 2018).

### **3.2 Le mythe du transhumanisme, nouvelle version du Phénix, du Juif errant, de la fontaine de Jouvence...**

Le Phénix est un aigle gigantesque et magnifique de l'ancienne Egypte, mais il ne peut se reproduire. Alors quand il se sent trop vieux, il construit son nid et il y met le feu. Mais des cendres de ce bûcher surgit alors le Phénix nouveau.

Le Juif errant est un cordonnier de Jérusalem qui, s'étant moqué du Christ portant sa croix, a été condamné à l'immortalité : il vieillit jusqu'à l'âge de 100 ans, il tombe alors malade, mais quand il guérit il a à nouveau trente ans. Au XV<sup>e</sup> siècle le grand chancelier de Florence l'a rencontré sous les traits de Giovanni Votaddio, au XVI<sup>e</sup> siècle on le retrouve en Allemagne sous les traits d'Ahasvérus.

La fontaine de Jouvence est une source qui restaure la jeunesse, il suffit d'en boire quelques gorgées. Jupiter a transformé la nymphe Jouvence en fontaine, mais où est donc cette source ? Alexandre le Grand l'a cherchée en Asie, l'empereur chinois Qin Shi ne l'ayant pas trouvée non plus il s'est fait enterrer pour l'éternité avec son armée de terre cuite, des druides l'auraient utilisée dans la forêt de Brocéliandre, des conquistadors l'ont cherchée vers la Floride...

Il y a fort à parier que les nouveaux entrepreneurs des *Biotech* dans le secteur appelé maintenant NBIC (nanotechnologies, biotechnologies, informatique et sciences cognitives) ont du être fascinés par ces mythes ancestraux sur la maîtrise de la mort. Citons Elon Musk (fondateur de SpaceX et de Telsa) qui vient de créer [Neuralink](#) pour développer des implants cérébraux connectés à des

ordinateurs ; citons Larry Page (fondateur de Google) qui vient de créer [Calico](#) pour chercher les codes génétiques expliquant la longévité humaine et son hérédité ; en France et plus modestement citons Laurent Alexandre, ancien médecin devenu entrepreneur (fondateur de Doctissimo, propriétaire de DNAVision) mais aussi auteur prolifique sur les thèses du transhumanisme (« certains d'entre vous dans cette salle vivront mille ans », conclusion de sa [conférence](#) TED 2012).

Pour le transhumanisme, une autre forme d'intelligence artificielle va émerger : la reproduction *in silico* du fonctionnement du cerveau va se combiner avec les modifications aujourd'hui possibles de l'ADN, pour créer le cyborg, cet être humain hybride. Dans cet imaginaire, la technologie est sans limite : on passe ainsi des réalités sur l'homme augmenté ou sur la thérapie génique, pour aller vers l'abolition des maladies et de la vieillesse par la techno-médecine et les nano-robots... et donc bientôt vers l'abolition de la mort ; l'apparition d'une post-humanité. Les scientifiques du domaine parlent d'imposture (Tritsch et Mariani, [Pour la Science](#) 2018), ce qui n'empêche pas le sérieux Beijing Genomics Institute, par exemple, d'autoriser le séquençage de milliers d'ADN humains pour chercher à identifier tous les « gènes de l'intelligence » ([Sciences Humaines](#) 2019).

#### **4. À qui profite ces discours de confusions ? Le nouveau pouvoir des empires numériques**

En combinant le mythe du transhumanisme (sur les objectifs à atteindre) et le mythe de la singularité (sur le calendrier à venir) nous avons les éléments pour décrypter tout ce discours brassant intentionnellement l'IA-*Informatique* avec l'IA-*Fiction* et cherchant surtout à donner le frisson. Mais il faut passer à un deuxième niveau : ce brasage et ces confusions sont au service de quels intérêts ?

Une première réponse consiste à dire que nous sommes un peu dans la même situation qu'en 1956 au séminaire de [Dartmouth](#), quand J. McCarthy et M. Minsky créent le néologisme ambigu d'Intelligence artificielle : il ne s'agirait que d'attirer l'attention du public pour financer des recherches originales. Cette réponse reste sans doute valable pour expliquer le financement d'une partie de l'écosystème IA, celui de certaines start-up qui cherchent à devenir « licornes » ([PbiFrance](#) 2019, [Guide des StartUps](#) 2019).

Mais aujourd'hui la réalité économique des empires numériques incite à défendre une autre thèse : ce frisson de peur et de rêve ne sert-il pas à détourner l'attention, pour finalement masquer une autre question, qui est celle d'un danger réel : la généralisation des applications d'IA aux mains des empires numériques menace progressivement toutes les institutions des états démocratiques. On pourrait dire sous forme de boutade « Attention en traversant les voies (médiatiques), un discours peut en cacher un autre ».

##### **4.1 L'IA de la Silicon Valley, dans le projet politique d'une société libertarienne**

Le projet politique des libertariens est de faire disparaître les institutions des états démocratiques, jugées vieilles et inefficaces, au profit de l'intérêt particulier érigé comme un principe : le droit absolu du contrôle sur soi-même (ce qui va de l'Open Source ou la dépénalisation des drogues... jusqu'à la liberté de disposer des armes à feu). Mais si, comme dans l'anarchisme, on s'attaque aux structures étatiques (impôts, monnaie, police, justice, éducation, santé... ) c'est ici pour mieux favoriser l'individualisme (et notamment l'accumulation capitaliste sur de nouveaux marchés, sans entraves, sans taxes, et sans redistribution ; le néolibéralisme n'est vu que comme un premier échelon).

On comprend alors que le projet de techno-planète de l'empire numérique des GAFAMI représente une étape inespérée dans le projet des libertariens, puisque cet empire rivalise avec les états pour assurer de plus en plus de tâches régaliennes : contester le monopole de la monnaie (Paypal, Bitcoin, Libra...), organiser la police préventive (PredPol) et la justice prédictive (LexMachina), réguler la liberté d'expression en filtrant les réseaux socio-numériques (Facebook), contester la souveraineté fiscale des états (Amazon, Google), casser le droit du travail (Uber), favoriser la manipulation des élections (Cambridge Analytica), organiser les systèmes de santé (IBM Watson) de mobilité (Google Car, Uber) de culture (Youtub, Netflix, bientôt Google avec les jeux vidéos en streaming), légiférer sur la presse (Google News et sa gestion des « droits voisins »), gérer le cadastre et l'urbanisme (Google Earth, blockchain cadastrale, Smart Cities), créer des universités (Cisco), maîtriser la conquête de l'espace (SpaceX)... et donc finalement gérer les grands choix de société à faire dans l'investissement mondial (à eux cinq les GAFAM ont une capitalisation boursière qui dépasse le PIB de *tous* les pays du monde, sauf USA, Chine et Japon, [Financepourtous](#) 2018).

Malgré toutes les contradictions de la Silicon Valley (soutien de l'Open source d'un coté mais exigence d'une hyper-protection des brevets de l'autre ; évasion fiscale d'un coté mais financements par la Darpa de l'autre ; multiples déclarations d'éthique d'un coté mais implication dans le programme Prism de l'autre...), on peut dire qu'elle assume totalement l'idéologie libertarienne soutenue par son grand gourou Peter Thiel fondateur de Paypal ([Usbeketrica](#) 2018). « Changer le monde » selon le slogan de Google offre certes un vernis plus respectable que de vouloir faire des profits, mais c'est bien plus qu'une simple formule publicitaire : il y a dix ans les idées du gourou ([Thiel](#) 2009) sur une course de la « bonne » technologie contre le « méchant » politique pour coloniser l'espace et les océans pouvaient encore nous faire sourire...

#### **4.2 L'IA des BHATX, dans le projet politique chinois du *Système de crédit social***

Les BHATX, Baidu Huawei Alibaba Tencent Xiaomi, jouent maintenant dans la même cour que les GAFAMI, aussi bien en terme de capitalisation boursière ([Liberation](#) 2018) que d'investissements dans l'IA ([Siecedigital](#) 2019). Ils sont ainsi devenus un des éléments importants du *soft power* dans la construction du capitalisme d'état en Chine, via par exemple par le développement des « Nouvelles routes *digitales* de la Soie » ([Mialhe](#) 2018) ou l'ambitieux parc technologique de l'IA à Pékin ([Dutton](#) 2018).

Il apparaît alors que le projet politique du Parti Communiste Chinois est de diriger le développement d'un hyper-capitalisme national dans une *société de contrôle*, harmonieuse en référence à la pensée confucéenne, en empêchant pour cela la naissance des institutions d'un état démocratique. D'un point de vue théorique, l'envahissement de la vie sociale par les technologies numériques peut en effet s'analyser comme une évolution de la *société disciplinaire* décrite par M. Foucault vers la *société de contrôle* décrite par G. Deleuze ([Ottavianni](#) 2003) et aujourd'hui la *société des traces* (Bouiller 2015).. D'un point de vue pratique, on peut effectivement observer une évolution, en Chine encore plus qu'ailleurs, depuis le *panoptique* disciplinaire imposé de l'extérieur (architecture des prisons, des classes d'écoles..) vers un *data-panoptisme* plus insidieux car il pénètre les comportements quotidiens : un classement des individus par une « *sousveillance* » constante des moindres faits et gestes de chacun, via tous les réseaux numériques et les omniprésents capteurs de données personnelles (Quessada 2010). Le *Système de Crédit Social* chinois, qui devrait permettre à partir de 2020 d'évaluer par des bons et mauvais points le degré de confiance que l'on peut avoir dans chaque individu et dans chaque entreprise, est aujourd'hui l'aboutissement impressionnant de cette société de contrôle. Sur ce sujet il apparaît vite que les médias mélangent facilement les peurs et les dangers, comme le montrent des auteurs qui ont une

bonne connaissance de la Chine, comme [S. Arsène](#) (2018) ou [P. Velghe](#) (2019).

On comprend en tout cas que les succès de l'empire numérique des BHATX représentent une étape inespérée pour la mise au point de ce *Système de Crédit Social* public, actuellement en phase de construction assez hétéroclite dans 43 grandes municipalités (document officiel du [Gouvernement](#) chinois en 2014 sur la « gestion sociale » et le « maintien de la stabilité » ; analyse de la [CNIL](#) en 2018). On peut en effet donner plusieurs exemples pour montrer que les BHATX assument totalement l'idéologie du capitalisme d'état :

- Alibaba fait (au minimum) profiter le gouvernement de son expérience. L'application privée de financement et de notation de crédit *Sésame Credit* a été lancée en 2015 par Ant Financial ([Les Echos](#) 2017) une filiale d'Alibaba « l'Amazon chinois » et de Didi Chuxing « l'Uber chinois », elle domine aujourd'hui le secteur banque et assurance. Mais son succès tient non seulement à sa connaissance fine des consommateurs (achats en ligne, paiements Alipay via mobiles, réseau social SocMe, site de rencontres Baihe) mais aussi à l'autorisation du gouvernement d'accéder aux bases publiques de données fiscales, juridiques et patrimoniales. Un algorithme confidentiel permet alors à *Sésame Credit* de donner à chacun des utilisateurs consentants un score de confiance qui va leur faciliter (ou pas) l'obtention d'un prêt ou d'une assurance, *mais aussi* l'utilisation prioritaire des transports, l'accès au soins sans dépôt de garantie ou même l'obtention d'un visa Schengen ou australien (et comme Alibaba partage votre score avec vos contacts et vice-versa... supprimer un de vos amis mal noté améliore votre propre score !). Or, même s'il est très difficile de savoir si toutes ces données sont systématiquement communiquées aux autorités publiques, il existe déjà des preuves formelles de la connexion depuis 2018 entre les bases de données de *Sésame Credit* et celles de la *Carte de citoyen* de la municipalité de Hangzhou : R. Raphael et Ling Xi ([Le Monde diplomatique](#) 2019) décrivent ainsi de multiples alliances « bricolées » entre l'état et les BHATX plutôt qu'un projet très centralisé de *Système de Crédit Social* public.

- Les BHATX sont officiellement impliqués dans la « gestion sociale » du gouvernement chinois. La Banque Populaire de Chine exploitait un service d'enquête sur les prêts personnels, avec ses propres bases de données sur les prêts et les cartes de crédit ; en 2015 elle a cédé cette activité à un groupe de huit sociétés pour gérer le contrôle des crédits aux PME et des crédits à la consommation. Qui mène ce groupe ? Tencent et Alibaba ([South China Morning Post](#) 2015). En 2017, le ministère des Sciences et Technologies a nommé une « Equipe Nationale pour l'IA ». Qui sont les premiers membres ? Baidu pour les voitures autonomes, Alibaba pour les villes intelligentes, Tencent pour les diagnostics médicaux, iFlytek pour la reconnaissance vocale et l'impressionnante licorne [SenseTime](#) pour la reconnaissance faciale ([South China Morning Post](#) 2017, [Les Echos](#) 2019). Le gouvernement chinois n'ambitionne pas seulement de noter chaque individu et de le filmer avec les grands programmes *Bouclier d'or* ([INA](#) 2018) puis *Oeil de lynx* ([Radio Free Asia](#) 2018) : 600 millions de caméras de surveillance sont prévus en 2020 (vidéo [Le Monde](#) 2017). Le gouvernement met aussi en place pour 2020 le système de *Crédit social Corporate* pour noter les entreprises. Les différentes bases de données ne sont ici pas encore unifiées, mais qui mène le consortium habilité ? Taiji Computer, Alibaba, Tencent, Huawei et VisionVera ([Le Monde](#) 2019).

## Conclusion

Nous nous sommes demandé s'il se cachait des logiques dangereuses derrière le discours médiatique de confusions entretenues entre IA-*Informatique*, IA-*Fiction* et Intelligence humaine. Nous avons répondu oui, mais nous avons trouvé deux logiques politiques très contradictoires : d'un

coté le droit absolu des libertariens au contrôle sur soi-même (qui sert les intérêts de l'oligopole nord-américain des GAFAMI, car il justifie de s'attaquer progressivement à toutes les institutions d'un état démocratique) et de l'autre coté le droit d'un capitalisme étatique au contrôle absolu sur ses citoyens (qui sert les intérêts de l'oligopole chinois des BHATX, car il justifie de s'associer à l'état pour organiser l'automatisation de la « gestion sociale » et bloquer toute apparition des institutions d'un état démocratique).

Dans une enquête policière on cherche d'abord à qui profite le crime et ensuite quelles sont les motivations. Si le crime était ici les confusions entretenues sur l'IA, nous avons proposé une réponse à la première question : au profit d'un duopole sino-américain de l'IA, car pendant que l'on disserte sur les peurs et les rêves de voir apparaître les cyborgs immortels, leurs affaires peuvent continuer à grandir en étant à l'abri de trop de débats sur leur puissance économique, sociale et politique.

Mais l'enquête doit progresser sur la question des motivations, car même si la vision des libertariens nord-américains et la vision du capitalisme d'état chinois partagent le même *solutionnisme technologique*, leurs contradictions semblent insurmontables :

- ces deux visions opposées du pouvoir ne sont-elles que concomitantes, auquel cas chacune continuera, pour le moment, à s'appuyer sur les technologies de l'IA pour réaliser ses objectifs ?
- ou bien doit-on tenter une analyse théorique de leur possible association ? Association qui nous mènerait alors vers une société post-démocratique, où nous serions *Seuls ensemble* selon la belle formule de la psychanalyste [S. Turkle](#) (2015) ; une société qui réussirait à allier la convivialité de technologies fortes et narcissiques (droit absolu de contrôle sur soi-même) en même temps que la police de relations sociales faibles et aliénantes (droit absolu de contrôle des citoyens) ?

## Bibliographie

### Publications académiques

- Arsène S. (2018), Trust in Ratings: China's Social Credit System, *Asian Perspectives*, may 2018 [Lien](#)
- Beaumat A. (2008). De la bêtise à la connerie.... *Empan*, 69(1), p. 163-165. [Lien](#)
- Benavent C. (2016), Big Data, algorithmes et marketing : rendre des comptes, *Statistique et société*, Vol. 4, N° 3, [Lien](#)
- Bengio Y. (2019), La révolution de l'apprentissage profond, *Interstices*, [Lien](#)
- Berthier T., Kempf O. (2017). Intelligence artificielle et conflictualité. Sur l'hypothèse de dérive malveillante d'une Intelligence Artificielle, *Revue de Gendarmerie nationale*, p. 149-157, [Lien](#)
- Boullier, D. (2015). Les sciences sociales face aux traces du big data: Société, opinion ou vibrations? *Revue française de science politique*, vol. 65(5), 805-828 [Lien](#)
- Boyd D., Crawford K. (2011), Six provocations au sujet du phénomène des Big Data, *Conférence Oxford Internet Institute*, [Lien](#)
- Cardon D. (2015), *À quoi rêvent les algorithmes ?* Seuil, 2015, [note de lecture](#)
- Casilli A. (2018), La plateforme comme mise au travail des usagers : Digital labor et nouvelles inégalités planétaires, in *Vers une République des Biens Communs ?*, Les Liens qui Libèrent, pp. 41-56, [Lien](#)
- Damasio A. (2010), *L'erreur de Descartes: la raison des émotions*, Odile Jacob, [Lien ebook](#)
- Desmurget M. (2019), *La Fabrique du crétin digital, les dangers des écrans pour nos enfants*, Seuil 2019, [Entretien](#)
- Dubuc B. (2019), Cours et conférences en neurosciences cognitives, *Université McGill*, [Lien](#)
- Ezratty O. (2016), *Les avancées de l'intelligence artificielle*, 159 p., eBook [Lien](#)

- Ezratty O. (2019), *Les avancées de l'intelligence artificielle*, 522 p., eBook [Lien](#)
- FNMR (2018) Fédération nationale des médecins radiologues, L'Intelligence Artificielle : rêve ou cauchemar du radiologue ? *Séminaire 2018*, [Lien](#)
- Ferber J. (1995), *Les systèmes multi-agents, vers une intelligence collective*, 513 p., InterEditions [Lien](#)
- Foucault M. (1978), La gouvernementalité, *Cours du Collège de France*, Dits et Ecrits Tome III texte n°239, [Lien](#)
- Ganascia J-G (2017), *Le mythe de la Singularité, faut-il craindre l'intelligence artificielle?* 144 p., Seuil, [Note de lecture](#)
- Ganascia J-G (2018), Le mythe de la singularité : faut-il craindre l'intelligence artificielle ? *Conférence Montpellier* [Lien](#)
- Gori R., Del Volgo M. (2009). L'idéologie de l'évaluation : un nouveau dispositif de servitude volontaire ? *Nouvelle revue de psychosociologie*, 8(2), 11-26
- Hadji C. (2018), Que doit-on craindre davantage : l'intelligence artificielle ou la bêtise humaine ? *The Conversation*, 4 janvier 2018, [Lien](#)
- LeCun Y. (2015), Deep learning, *Conférence USI* [Lien](#)
- LeCun Y. (2016), L'apprentissage profond : une révolution en intelligence artificielle. *Leçon inaugurale au Collège de France*. [Lien](#) et [Diaporama](#)
- Lemoine P. (2014), *La nouvelle grammaire du succès, la transformation numérique de l'économie française*, Rapport au Gouvernement, 328 p., [Lien](#)
- Marosan M. (2019), Le devenir robot de l'humain, *The Conversation*.15 janvier 2019 [Lien](#)
- Newell A., Simon H. (1963), Computers in psychology, in *Handbook of Mathematical Psychology*. John Wiley, [Lien](#)
- Ottaviani D. (2003), Foucault - Deleuze : de la discipline au contrôle, in *Lectures de Michel Foucault*, ENS Editions, p. 59-73 [Lien](#)
- Paquet P. (1997), L'utilisation des réseaux de neurones artificiels en finance, *Working paper*, 26 pages, IAE Orléans, [Lien](#)
- Quessada D. (2010). De la sousveillance: La surveillance globale, un nouveau mode de gouvernementalité. *Multitudes*, 40(1), 54-59. [Lien](#)
- Romero M., Lille B., Patiño A. (2017), *Usages créatifs du numérique pour l'apprentissage au XXIe siècle*, Presses de l'Université du Québec, 186 p., [Lien](#)
- Rouvray A. (2012), Face à la gouvernementalité algorithmique, repenser le sujet comme puissance, *Working paper*, Pré-publication, [Lien](#)
- Sadin E. (2009), *Surveillance globale: Enquête sur les nouvelles formes de contrôle*, Flammarion, [Lien](#)
- Scaruffi P. (2016), *Intelligence is not artificial*, 298 p., [Lien](#)
- Turing A. (1950) Computing Machinery and Intelligence. *Mind* 49: 433-460. [Lien](#)
- Turkle S. (2015), *Seuls ensemble. De plus en plus de technologies de moins en moins de relations humaines*, L'Échappée, [Note de lecture](#)
- Velghe P. (2019), Lire la Chine : Internet des Objets, surveillance et gestion sociale en RPC, *Perspectives chinoises*, 2019-1 [Lien](#)
- Zazzo R. (1983), *Où en est la psychologie de l'enfant ?* Denoël, Folio, [Note de lecture](#)
- Zuiderveen-Borgesius Z. (2018), *Discrimination, intelligence artificielle et décisions algorithmiques*, Rapport Conseil de l'Europe, 53 p. [Lien](#)

#### **Publications d'actualité** (dans l'ordre où elles apparaissent dans le texte)

- [Interstices](#) (2016) H. Prade : Marvin Minsky, un père visionnaire de l'intelligence artificielle
- [Will-Bridge](#) (2019), L'intelligence artificielle et le bridge
- [Bipame](#) (2019), Intelligence artificielle : État de l'art et perspectives pour la France, 40 pages
- [Youtube](#) (2018), Google's AI Assistant Can Now Make Real Phone Calls You
- [MIT Technology Review](#) (2019), Daniel Zender : Artificial Intelligence, We analyzed 16,625 papers to figure out where AI is headed next



[Youtube](#) (2010), Cours de E. Pruvost : Le perceptron, premier réseau doté d'apprentissage

[Smile](#) (2018), Livre blanc Deep Learning, 64 pages

[Microsoft Research](#) (2015), Deep Residual Learning

[OMPI](#) (2019), Rapport Tendances technologiques 2019 – Intelligence artificielle

[Opinions libres](#) (2018), O. Ezratty : Que devient l'IA symbolique ?

[InternetActu](#) (2017), R. Sussan : Les limites du deep learning et comment les dépasser

[L'UsineNouvelle](#) (2018), A la conférence FranceisAI, le deep learning montre ses faiblesses

[Medium](#) (2017), La nouvelle performance d'AlphaGo est DINGUE

[Znet](#) (2017), AlphaGo, l'IA prend sa retraite de joueur après une ultime victoire

[Le Figaro](#) (2019), Surdouée au jeu de Go, l'IA de Deepmind rate un contrôle de mathématiques de lycée

[Harvard Business Review](#) (2016), How Artificial Intelligence Will Redefine Management

[Mediapart](#) (2018), L'enfer du «social ranking»: quand votre vie dépend de la façon dont l'Etat vous note

[Gizmodo](#) (2013) This classical music was created by a supercomputer in less than a second

[Huffingtonpost](#) (2018) Ce tableau a été créé par une intelligence artificielle, mais est-ce de l'art?

[Actualité](#) (2018), The Road, un voyage à la Kerouac écrit par une intelligence artificielle

[BBC News](#) (2018), Fake Obama created using AI video tool

[Wikipedia](#) (2019), Lsjbot, un programme de création automatique d'articles Wikipédia

[L'UsineDigitale](#) (2019), Cortana peut maintenant tenir une conversation pendant plusieurs minutes sans en perdre le fil

[Opinions libres](#) (2019), Où en est l'IA émotionnelle ?

[Wintics](#) (2018), Quand la Data Science devient créative avec les GAN

[Interstices](#) (2016), La conscience d'une machine

[PourLaScience](#) (2019), Plusieurs milliers d'espèces de plancton repérées par Tara Oceans sont inconnues

[LaRecherche](#) (2016), Les ondes gravitationnelles ont été détectées

[Université de Montréal](#) (2017), la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle

[InternetActu](#) (2019), Intelligence artificielle, des limites de l'éthique aux promesses de la régulation

[Opinions libres](#) (2019), O. Ozeratty : Cerebras, le plus grand processeur d'IA

[Conférence TED](#) (2009), Ray Kurzweil: Une université pour la singularité à venir

[Opinions libres](#) (2019), O. Ozeratty : Ce que prépare Neuralink

[GenerationCare](#) (2017), Que sait-on de Calico, le laboratoire d'Alphabet qui lutte contre le vieillissement ?

[Conférence TED](#) (2012), L. Alexandre : le recul de la mort, l'immortalité à brève échéance?

[Pour la Science](#) (2018), D. Tritsch, J Mariani : L'imposture du transhumanisme

[Sciences Humaines](#) (2019), J-F Bouvet : Les apprentis sorciers de l'eugénisme

[PbiFrance](#) (2019), Les start-up françaises de l'intelligence artificielle se multiplient

[Opinions libres](#) (2019), O. Ozeratty : Guide des Startups 2019

[Financepourtous](#) (2018), GAFA, GAFAM ou NATU : les nouveaux maîtres du monde

[Usbeketrica](#) (2018), Peter Thiel, l'homme qui voulait achever la démocratie

[CatoUnbund](#) (2009), Peter Thiel : The Education of a Libertarian

[Liberation](#) (2018), BATX : les quatre géants du web chinois

[Medium](#) (2018), T. Dutton : An Overview of Artificial Intelligence Strategies)

[Gouvernement chinois](#) (2014), Planning Outline for the Construction of a Social Credit System (2014-2020)

[CNIL](#) (2018), R. Chatellier : Le Crédit Social chinois et le dilemme éthique de la confiance par la notation

[Les Echos](#) (2017), Comment Ant Financial a révolutionné la finance chinoise

[Le Monde diplomatique](#) (2019), R. Raphael, Ling Xi : Quand l'État organise la notation de ses citoyens, Bons et mauvais Chinois

[South China Morning Post](#) (2015), PBOC calls on Alibaba, Tencent to help develop credit reporting market

[South China Morning Post](#) (2017), China recruits Baidu, Alibaba and Tencent to AI 'national team'

[Les Echos](#) (2019), En Chine, la vie sous l'oeil inquisiteur des caméras

[INA](#) (2018), Cybercontrôle en Chine : l'omni-surveillance à l'ère du numérique

[Radio Free Asia](#) (2018), China Aims For Near-Total Surveillance, Including in People's Homes

[Le Monde](#) (2017), En Chine, des caméras devinent qui sont les passants dans la rue

[Le Monde](#) (2019), La Chine met en place un dispositif pour mieux contrôler les sociétés étrangères