



Les avancées de l'intelligence artificielle – 8

Voici la suite de cette longue série d'articles sur les avancées de l'intelligence artificielle. Après **un peu de sémantique**, **un tour des techniques de l'IA**, **l'étude de cas d'IBM Watson**, **un tour d'horizon des startups américaines de l'IA**, puis de celles qui sont **acquises par les grands groupes**, **les startups françaises du secteur**, et l'état de l'art de la **connaissance du cerveau**, attaquons nous aux évolutions technologiques qui pourraient permettre à l'IA de faire des sauts quantiques.

L'IA a connu des vagues diverses d'hivers et de renaissances. Pour certains, il s'agit plutôt de vaguelettes. Les récentes "victoires" de l'IA comme dans Jeopardy (2011) et AlphaGo (2016) donnent l'impression que des sauts quantiques ont été franchis. C'est en partie une vue de l'esprit car ces progrès sont sommes toutes modestes et réalisés dans des domaines très spécialisés, surtout pour le jeu de Go.

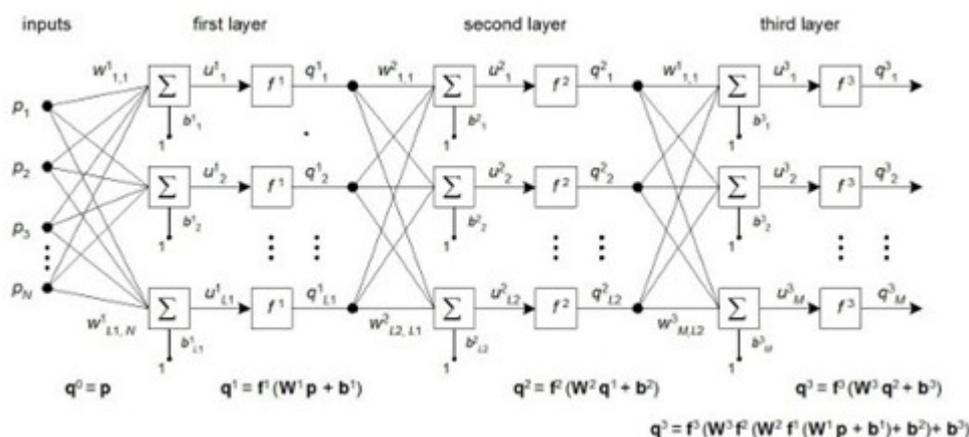
Peut-on décortiquer par quels biais les progrès dans l'IA vont s'accélérer ? Nous avons vu dans les parties précédentes qu'il était difficile de faire la part des choses entre avancées liées à l'immatériel et celles qui dépendent du matériel. Je vais commencer par les algorithmes et logiciels puis aborder la partie matérielle. Avec en interlude, un passage sur l'application de la loi de Moore dans la vraie vie qui est bien différente des belles exponentielles présentées à tout va !

Algorithmes et logiciels

Nous verrons plus loin que le matériel continuera de progresser, même si c'est un chemin semé d'embûches du côté des processeurs.

S'il y a bien une loi de Moore difficile à évaluer, c'est celle des algorithmes et logiciels ! Personne ne la mesure et pourtant, une bonne part des progrès numériques vient de là et pas seulement de l'augmentation de la puissance du matériel.

Les réseaux neuronaux à boucle de feedback et le deep learning auto-apprenants sont maintenant anciens et leur progression est lente dans le principe. Leur mise en œuvre s'améliore beaucoup grâce aux possibilités matérielles qui permettent de créer des réseaux neuronaux multicouches allant jusqu'à 14 couches.



A chaque fois qu'un record est battu comme avec AlphaGo, il résulte de la combinaison de la force du matériel, du stockage et du logiciel. Qui plus est, ces records de l'IA portent sur des domaines très spécialisés. La variété et les subtilités des raisonnements humains sont encore loin. Mais elles ne sont pas hors de portée. Notre cerveau est une machine hyper-complexe, mais ce n'est qu'une machine donc potentiellement imitable.

La recherche progresse en parallèle dans les techniques de reconnaissance d'images (à base de réseaux de neurones et de machine learning), de la parole (itou) et de l'analyse de données (idem). Les algorithmes génétiques sont de leur côté utilisés pour trouver des chemins optimums vers des solutions à des problèmes complexes intégrant de nombreux paramètres, comme pour trouver le chemin optimum du voyageur du commerce.

C'est dans le domaine de l'**intelligence artificielle intégrative** que des progrès significatifs peuvent être réalisés. Elle consiste à associer différentes méthodes et techniques pour résoudre des problèmes complexes voire même résoudre des problèmes génériques. On la retrouve mise en œuvre dans les agents conversationnels tels que ceux que permet de créer IBM Watson ou ses concurrents.

Dans le jargon de l'innovation, on appelle cela de l'innovation par l'intégration. C'est d'ailleurs la forme la plus courante d'innovation et l'IA ne devrait pas y échapper. Cette innovation par l'intégration est d'autant plus pertinente que les solutions d'IA relèvent encore souvent de l'artisanat et nécessitent beaucoup d'expérimentation et d'ajustements.

Cette intégration est un savoir nouveau à forte valeur ajoutée, au-delà de l'intégration traditionnelle de logiciels via des APIs classiques. Cette intelligence artificielle intégrative est à l'œuvre dans un grand nombre de startups du secteur. Le mélange des genres n'est pas évident à décrypter pour le profane : machine learning, deep learning, support vector machines, modèles de Markov, réseaux bayésiens, réseaux neuronaux, méthodes d'apprentissage supervisées ou non supervisées, etc. D'où un discipline qui est difficile à benchmarker d'un point de vue strictement technique et d'égal à égal. Ce d'autant plus que le marché étant très fragmenté, il y a peu de points de comparaison possibles entre solutions. Soit il s'agit de produits finis du grand public comme la reconnaissance d'images ou vocale, et d'agents conversationnels très à la mode en ce moment, soit il s'agit de solutions d'entreprises exploitant des jeux de données non publics. Un nouveau savoir est à créer : le benchmark de solutions d'IA ! Voilà un métier du futur !

La **vie artificielle** est un autre pan de recherche important connexe aux recherches sur l'IA. Il s'agit de créer des modèles permettant de simuler la vie avec un niveau d'abstraction plus ou moins élevé. On peut ainsi simuler des comportements complexes intégrant des systèmes qui s'auto-organisent, s'auto-réparent, s'auto-répliquent et évoluent d'eux-mêmes en fonction de contraintes environnementales.

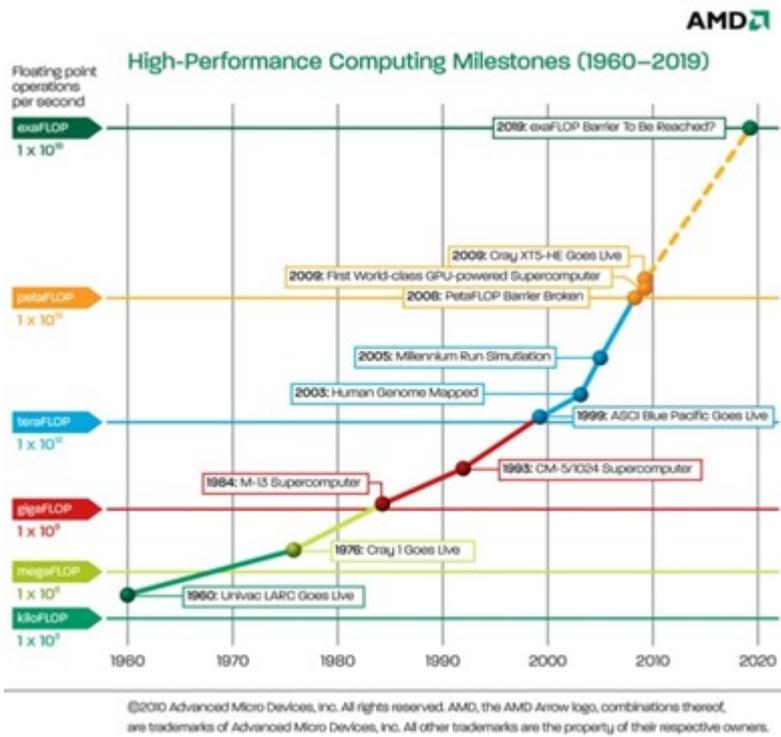
Jusqu'à présent, les solutions d'IA fonctionnaient à un niveau de raisonnement relativement bas. Il reste à créer des machines capables de gérer le sens commun, une forme d'intelligence génétique capable à la fois de brasser le vaste univers des connaissances – au-delà de nos capacités – et d'y appliquer un raisonnement permettant d'identifier non pas des solutions mais des problèmes à résoudre. Il reste à apprendre aux solutions d'IA d'avoir envie de faire quelque chose. On ne sait pas non plus aider une solution d'IA à prendre du recul, à changer de mode de raisonnement dynamiquement, à mettre plusieurs informations en contexte, à trouver des patterns de ressemblance entre corpus d'idées d'univers différents permettant de résoudre des problèmes par analogie. Il reste aussi à développer des solutions d'IA capables de créer des théories et de les vérifier ensuite par l'expérimentation.

Pour ce qui est de l'ajout de ce qui fait de nous des êtres humains, comme la sensation de faim, de peur ou d'envie, d'empathie, de besoin de relations sociales, on en est encore loin. Qui plus est, ce n'est pas forcément nécessaire pour résoudre des problèmes courants de l'univers des entreprises. Comme l'indique si bien **Yuval Noah Harari**, l'auteur du best-seller "Sapiens" qui interviendra en juin dans la **conférence USI** organisée par Octo Technology à Paris, "*L'économie a besoin d'intelligence, pas de conscience*" ! Laissons donc une partie de notre intelligence voire une intelligence plus développée aux machines et conservons la conscience, les émotions et la créativité !

La loi de Moore dans la vraie vie

La loi de Moore est la pierre angulaire de nombreuses prédictions technologiques, notamment pour ce qui concerne celles de l'intelligence artificielle. Présentée comme immuable et quasi-éternelle, cette loi empirique indique que la densité des transistors dans les processeurs double tous les 18 à 24 mois selon les versions. Elle est aussi déclinée à foison pour décrire et prédire divers progrès techniques ou technico-économiques. Cela peut concerner la vitesse des réseaux, la capacité de stockage, le coût d'une cellule solaire photovoltaïque ou celui du séquençage d'un génome humain. Une progression n'en entraîne pas forcément une autre. Le coût peut baisser mais pas la performance brute, comme pour les cellules solaires PV. On peut donc facilement jouer avec les chiffres.

La loi de Moore est censée s'appliquer à des solutions commercialement disponibles, et si possible, en volume. Or ce n'est pas toujours le cas. Ainsi, l'évolution de la puissance des supercalculateurs est mise en avant comme un progrès technique validant la loi de Moore. Or, ces calculateurs sont créés avec des moyens financiers quasiment illimités et n'existent qu'en un seul exemplaire, souvent réalisé pour de la recherche militaro-industrielle ou de grands projets de recherche (aérospatial, génomique, météo). Ce que l'on peut observer dans la belle exponentielle ci-dessous issue d'AMD.



Dans la plupart des cas, ces technologies “de luxe” se retrouvent dans des produits grand public après quelques années. Ainsi, la puissance des super-calculateurs des années 1990 s’est retrouvée dans les consoles de jeu des années 2000. Au lieu de faire des calculs en éléments finis pour des prévisions météo, les consoles de jeux calculent des millions de polygones pour simuler des images en 3D temps réel. Mais cette puissance n’est pas homothétique dans toutes les dimensions. Si la puissance de calcul est similaire, les capacités de stockage ne sont pas les mêmes.

Examinons donc de près comment cette fameuse loi s’applique pour des objets numériques grand public. Prenons trois cas d’usages courants : un **laptop** plutôt haut de gamme en 2006 et en 2016, l’évolution de l’**iPhone** entre sa première édition lancée en juin 2007 et l’iPhone 6S lancé en septembre 2015 et puis l’évolution du **haut débit fixe** sur 10 ans.

En appliquant une belle loi de Moore uniforme, les caractéristiques techniques de ces trois larrons devraient doubler tous les deux ans au minimum. Sur une période de 10 ans, cela donnerait 2 puissance 5 soient **x32** et sur 8 ans, **x16**. Si le doublement intervenait tous les 18 mois, ces facteurs seraient respectivement de **x101** et **x40**.

Commençons par un **laptop** haut de gamme à prix équivalent entre 2006 et 2016. J’ai comparé deux modèles plutôt haut de gamme de la même marque : un Asus W7J de 2006 et un Asus Zenbook UX303UA de 2016, certes sorti en 2015. Mais entre fin 2015 et mi 2016, il n’y a pas eu de changements d’architecture des laptops, qui collent à la roadmap d’Intel.

Aucun paramètre technique n’a évolué d’un facteur x32 et à fortiori d’un facteur x100. Ceux qui ont le mieux progressé et qui ont un impact sur la performance perçue par l’utilisateur sont la vitesse du moteur graphique (x12) et celle du Wi-Fi (x24). Pour le reste, les gains sont très modestes. Le processeur est “seulement” 3,7 fois plus rapide. La résolution des écrans a augmenté mais la résolution limitée de l’œil rend caduque cette progression dès lors qu’un écran atteint la résolution 4K, qui commence à apparaître sur certains laptops.

Asus W7J 2006



Asus Zenbook UX303UA 2016



Le plus grand retardataire est la batterie qui n'évolue quasiment pas. L'autonomie des laptops a progressé non pas grâce aux batteries mais à la baisse de consommation des processeurs et autres composants électroniques ainsi qu'à l'intelligence intégrée dans les systèmes d'exploitation, aussi bien Windows que MacOS. Les derniers processeurs Intel savent éteindre certaines de leurs parties lorsqu'elles ne sont pas utilisées. Par contre, la densité des batteries s'est un peu améliorée et leur cure d'amaigrissement a permis de créer des laptops plus fins.

	2006	2016	Multiple de performance	Commentaire
PC laptops servant à la comparaison	Asus W7J	Asus Zenbook UX303UA		Deux laptops plutôt haut de gamme au moment de l'évaluation.
CPU	Intel Core 2 Duo T7200	Intel Core i7 6500		Processeurs mobiles haut de gamme de laptop pour ces deux périodes.
Cœurs	2	2		1 Stabilité du nombre de cœurs. Elle augmente plutôt sur les CPU de desktops.
Clock	2 GHz	2,5 GHz	1,25	Faible évolution de l'évolution de la vitesse d'horloge.
Gravure et évolution de la densité	65 nm	14 nm	21,6	Le nombre de transistors augmente, notamment du fait du GPU.
Transistors	291 millions	1,5 milliards (estimation)	5,2	Intel ne communique plus sur le nombre de transistors par chipset depuis 2015.
Cache	4 Mo	4 Mo		1 Rien de changé.
Performance en Passmark	1175	4328	3,7	Mesure la plus importante de performance, perçue par l'utilisateur.
Consommation (TDP)	34W	7.5 à 25 W	4,5	Baisse de la consommation qui permet d'augmenter l'autonomie.
Mémoire RAM	1,5 Go	8 Go	5,3	Contribue à améliorer la vitesse de passage d'une application à l'autre.
Fréquence mémoire RAM	533 MHz	1600 MHz	3	Permet de transférer plus rapidement les données en mémoire.
GPU	GeForce 7400 128 Mo	Intel HD Graphics 520		Processeur graphique.
G3D Benchmark	63	768	12,2	Belle évolution sachant que le GPU intégré aux chipsets Intel n'est pas ce qu'il y a de mieux pour un laptop vs une carte nVidia ou AMD.
Stockage	120 Go HDD 5400 rpm	512 Go SSD	4,2	Augmentation modérée du fait du passage du HDD au SSD. Avec un HDD de 2 To, le ratio serait de 17.
Bus stockage	SATA 1 (150)	SATA 3 (600)	4	6 Gbits/s sachant qu'avec un port M.2, on peut aller jusqu'à 1,6 Go/s.
Ethernet	1 Gbits/s	1 Gbits/s		1 Rien de changé.
Wi-Fi	802.11 g à 54 Mbits/s	802.11 ac à 1300 Mbits/s	24	La plus grande évolution dans un laptop !
Résolution écran	1 366x768	3200x1800	5,49	Belle évolution mais qui sera limitée à terme par la résolution oculaire.
Résolution webcam	1,3 Mpixels	720p		Pas d'évolution.
Batterie	49 Wh	56 Wh	1,14	Les batteries évoluent peu. La consommation baisse, mais pas l'énergie disponible.
Poids	1,95 Kg	1,3 Kg	1,50	Ca s'allège doucement.
Epaisseur	36,9 mm	19,2 mm	1,92	Ca maigrit bien. L'impact du MacBook Air lancé en 2010.
Prix	1 800 €	1 560 €	1,15	Prix relativement stable.

Du côté de l'**iPhone**, la situation est plus contrastée et bien meilleure que pour les laptops. Deux dimensions techniques ont bien progressé : le processeur qui est x18 fois plus rapide et la communication data Internet mobile qui est **x781** fois plus rapide, tout du moins en théorie, car d'un point de vue pratique, le ratio réel est plus raisonnable. Et encore, c'est lié au choix peu hardi du Edge au moment de la sortie du premier iPhone alors que la 3G aurait très bien pu être supportée.

iPhone "1" juin 2007



iPhone 6S septembre 2015



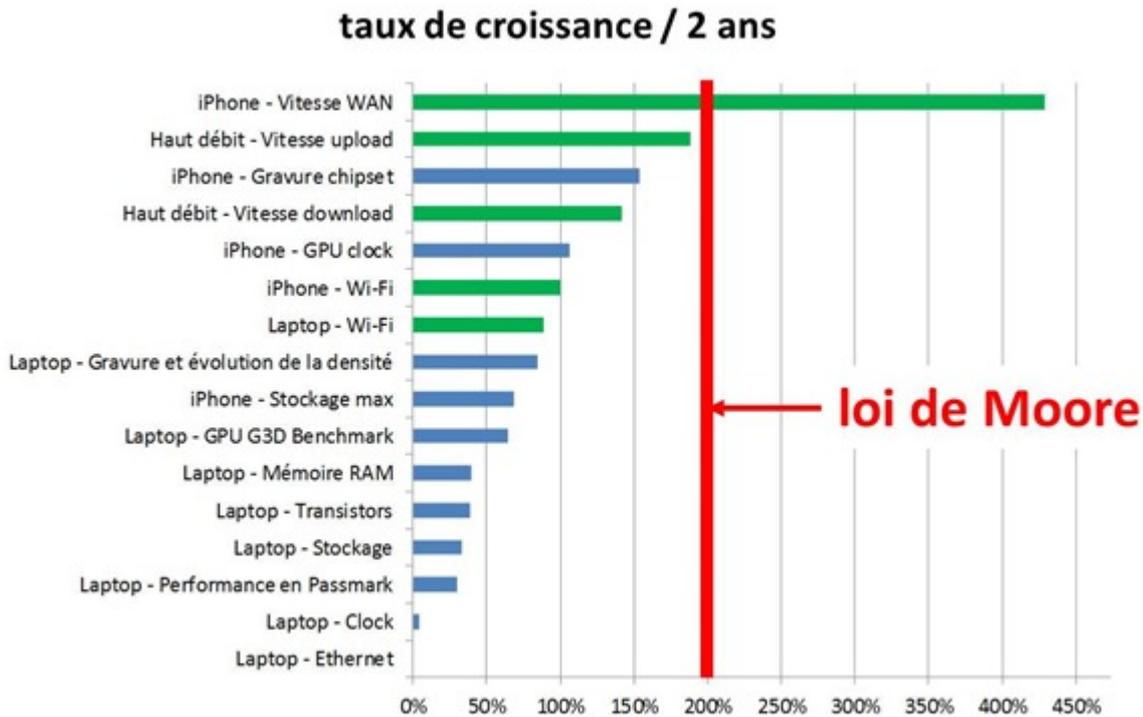
Contrairement aux laptops, au lieu de voir les prix baisser, ils augmentent, positionnement haut de gamme d'Apple oblige. Le poids augmente aussi car l'iPhone 6S a un écran plus grand que celui du premier iPhone. Et comme pour les laptops, la capacité de la batterie a très peu augmenté. J'ai indiqué les résolutions d'écran et de capteurs vidéo sachant qu'il n'y a pas de raison objective de vouloir poursuivre ad-vitam la loi de Moore pour ce qui les concerne.

	Jun 2007	Septembre 2015	Multiple de performance	Commentaire
	iPhone 1	iPhone 6S		
CPU	Samsung S5L8900	Apple A9		Les processeurs des iPhone sont conçus par Apple et fabriqués par Samsung et TSMC.
Cœurs et clock	un cœur 32 bits à 412 MHz	double cœur, 64 bits 1,8 GHz	18	Meilleure évolution de la clock grâce aux noyaux ARM basse consommation.
Gravure	90 nm	14 nm	41,3	
GPU	PowerVR MBX Lite 60 MHz	PowerVR GT7600		
Stockage max	16 Go	128 Go	8	Progression des mémoires Flash continue.
Wifi	Wi-fi g à 54 Mbits/s	Wi-fi ac à 866 Mbits/s	16	Progression moindre que sur PC car le ac s'applique à géométrie variable.
WAN	Edge à 384 kbit/s	LTE à 300 Mbits/s	781	La plus forte évolution technologique inventoriée ici !
Ecran	480 x 320	1334*750	6,5	Meilleure progression avec le 6 Plus et aussi Sony Xperia Z5 en 4K.
Capteur dorsal	2 Mpixels	12 Mpixels	6	Certains smartphones atteignent 20 Mpixels mais cela ne sert à rien.
Batterie	1400 mAh	1642 mAh	1,2	Très faible évolution. Le 6 Plus a une batterie de 2900 mAh.
Poids	135 g	143 g	0,94	Le 6S est plus lourd car l'écran est plus grand.
Prix	659 € pour 16 Go	1079 € pour 128 Go 859 € pour 16 Go	0,6	Régression, eu égard à la politique de prix et de marge d'Apple.

La situation est assez différente du côté du **haut débit fixe**. Vous pouvez stagner pendant une décennie à la même vitesse d'accès à Internet et bénéficier tout d'un coup d'un progrès soudain appliquant 10 ans de loi de Moore. Si vous passez par exemple d'un ADSL à 12 Mbits/s en download et 1 Mbits/s en upload à de la fibre chez Free à 1 Gbits/s en download et 200 Mbits/s en upload, le facteur multiplicateur est respectivement de x83 et x200. Si vous partiez d'un débit encore plus faible du fait d'un plus grand éloignement des centraux télécoms, le facteur multiplicateur serait encore plus élevé. Mais plus votre débit ADSL d'origine est faible, plus faibles sont les chances de voir la fibre arriver chez vous du fait des travaux d'infrastructure à réaliser pour passer les fourreaux transportant la fibre du central télécom jusqu'à chez vous !

Chez les autres opérateurs que Free, le facteur multiplicateur dépend de la technologie utilisée. Chez Numericable, c'est du FTTB à la performance à géométrie variable selon l'âge du capitaine et surtout un débit montant assez limité. Chez Orange, vous avez des taquets de débits à 100, 200 et 500 Mbits/s en download et de 50 Mbits/s à 200 Mbits/s en upload selon l'offre commerciale. Et si vous attendez toujours la fibre, la loi de Moore vous concernant est un encéphalogramme tout plat !

En ne conservant que les paramètres techniques où la loi de Moore est pertinente, voici donc ce que cela donne sous une autre forme, à savoir la progression moyenne tous les deux ans. On voit qu'à part la data WAN, on est loin du doublement tous les deux ans de la performance !

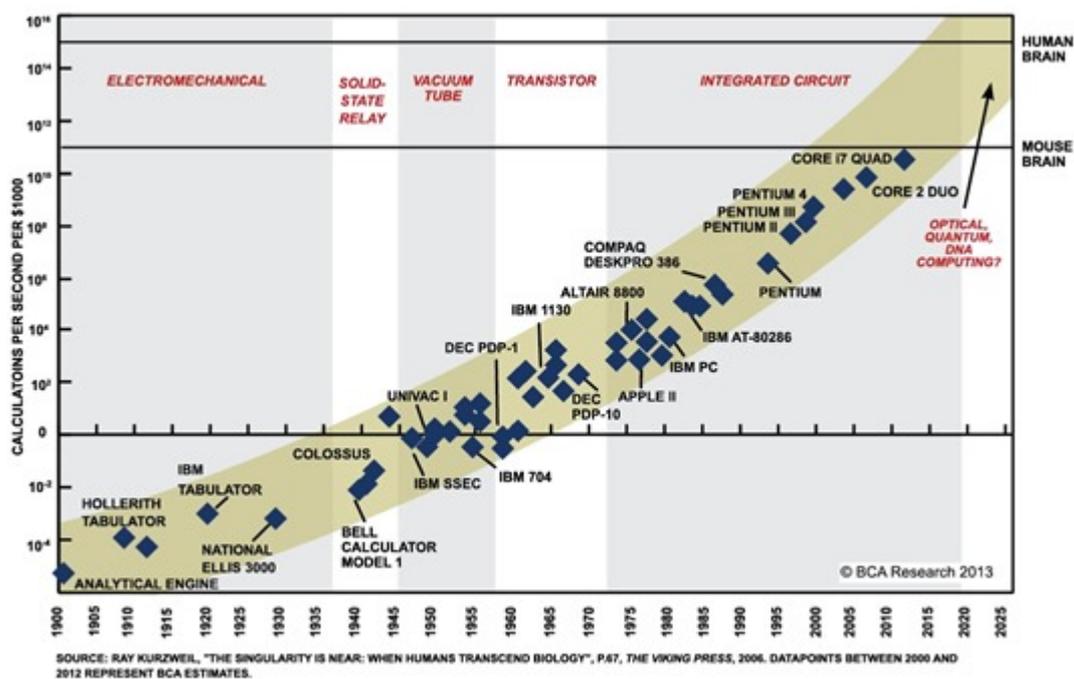


La loi de Moore s'applique bien mieux aux liaisons réseaux haut débit fixe et mobiles qu'à la capacité de calcul et de stockage, surtout sur ordinateurs personnels. Cela explique indirectement la montée en puissance des architectures en cloud. On peut plus facilement répartir une grosse charge de calcul sur des serveurs que sur des postes de travail ou des mobiles. On retrouve cette architecture dans Siri qui traite une bonne part de la reconnaissance vocale côté serveurs. Au passage, la loi de Moore de la vraie vie valide aussi le scénario de fiction de "Skynet" des films Terminator où c'est une intelligence logicielle distribuée sur des millions de machines dans le réseau qui provoque une guerre nucléaire !

Alors, la loi de Moore est foutue ? Pas si vite ! Elle avance par hoquets. Il reste encore beaucoup de mou sous la pédale pour faire avancer la puissance du matériel et sur lequel l'IA pourrait surfer.

Puissance de calcul

La fameuse loi de Moore est mise en avant par les singularistes pour prédire le dépassement de l'homme par l'IA à une échéance de quelques décennies. Seulement voilà, la validation dans la durée de cette loi empirique de Moore n'est pas triviale comme nous venons de le voir.



La question est revenue au-devant de la scène alors que cette loi fêtait ses 50 ans d'existence. Un anniversaire commenté pour annoncer la fin de ses effets, tout du moins dans le silicium et les technologies CMOS. Cette technologie est sur le point d'atteindre un taquet aux alentours de 5 nm d'intégration sachant que l'on est déjà à 10 nm à ce jour, notamment chez Intel, et à 14 nm en version commerciale (Core M et Core i de génération Skylake 2015). Les architectures multi-cœurs atteignent de leur côté leurs limites car les systèmes d'exploitation et les applications sont difficiles à ventiler automatiquement sur un nombre élevé de cœurs, au-delà de 4.

Le schéma ci-dessus et qui vient de Kurzweil n'a pas été mis à jour depuis 2006. Il est difficile d'obtenir un schéma sur l'application de la loi de Moore au-delà de 2010 pour les processeurs. Est-ce parce que l'évolution de la puissance de calcul s'est calmée depuis ? Dans le même temps, les découvertes en neuro-biologies évoquées dans l'article précédent augmentent de plusieurs ordres de grandeur la complexité de la modélisation du fonctionnement d'un cerveau humain. Bref, cela retarde quelque peu l'échéance de la singularité.



L'excellent dossier After Moore's Law, paru dans The Economist en mars 2016, détaille bien la question en expliquant pourquoi la loi de Moore des transistors CMOS pourrait s'arrêter d'ici une douzaine d'année lorsque l'on descendra au niveau des 5 nm d'intégration. Et encore, la messe n'est pas encore dite. A chaque nouvelle

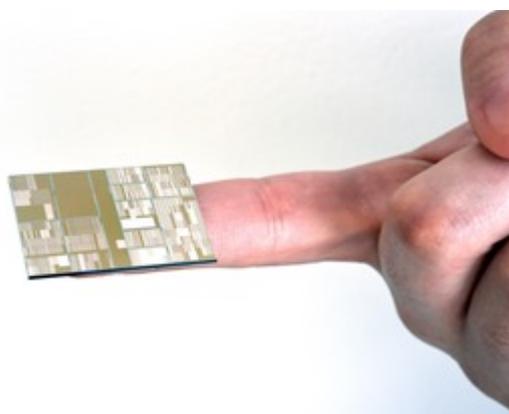
génération d'intégration, les fondateurs se demandent s'ils vont pouvoir faire descendre réellement le coût de fabrication des transistors. En-dessous de 14 nm, ce n'est pas du tout évident. Mais l'ingénierie humaine a des ressources insoupçonnables comme elle l'a démontré dans les générations précédentes de processeurs CMOS !

Il faudra tout de même trouver autre chose, et en suivant divers chemins de traverse différents des processeurs en technologie CMOS.

Voici les principales pistes connues à ce jour et qui relèvent toutes plutôt du long terme :

Continuer à descendre coûte que coûte le niveau d'intégration

En 2015, IBM et Global Foundries créaient une première en testant la création d'un processeur en **technologie 7 nm** à base de silicium et de germanium, battant le record d'Intel qui est à ce jour descendu à 10 nm. L'enjeu clé est de descendre en intégration sans que les prix n'explorent. Or, la gravure en extrême ultra-violet qui est nécessaire pour "dessiner" les transistors sur le silicium est complexe à mettre au point et plutôt chère.

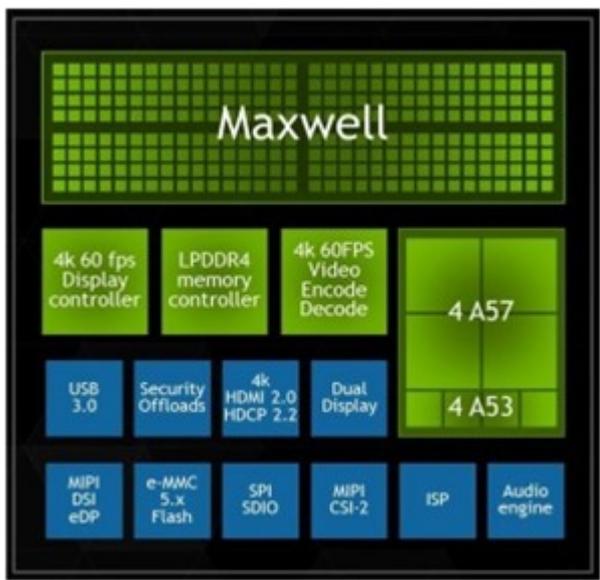


Le multi-patterning, que j'explique [ici](#), permet d'en contourner les limitations. Mais il coûte tout aussi cher car il ajoute de nombreuses étapes à la fabrication des chipsets et peut augmenter le taux de rebus. La loi de Moore s'exprime en densité de transistors et aussi en prix par transistors. Si la densité augmente mais que le prix par transistor augmente aussi, cela ne colle pas pour les applications les plus courantes.

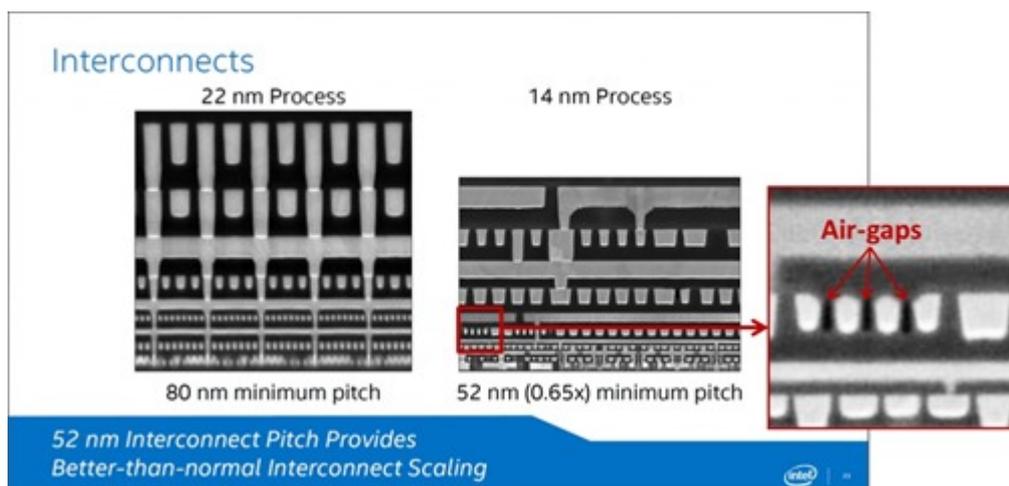
Créer des processeurs spécialisés

Ils sont notamment utiles pour créer des réseaux neuronaux, comme nous l'avons déjà vu dans la **seconde partie** de cette série. La piste est intéressante et est déjà très largement utilisée dans le cadre des GPU ou des codecs vidéo qui sont souvent décodés dans le matériel et pas par logiciel, comme le format HEVC qui est utilisé dans la diffusion de vidéo en Ultra Haute Définition (4K).

C'est l'approche de Nvidia avec ses chipsets X1 (*ci-dessous*) à 256 cœurs ou plus, qui sont utilisés dans la reconnaissance d'images des véhicules autonomes ou à conduite assistée comme les Tesla S. Ces GPU simulent des réseaux neuronaux avec une faculté d'auto-apprentissage. La piste se heurte cependant aux limites de la connectique. Pour l'instant, dans les réseaux neuronaux matériels, chaque neurone n'est relié qu'à ceux qui sont avoisinants dans leur plan. Dans le cerveau, l'intégration des neurones est tridimensionnelle.



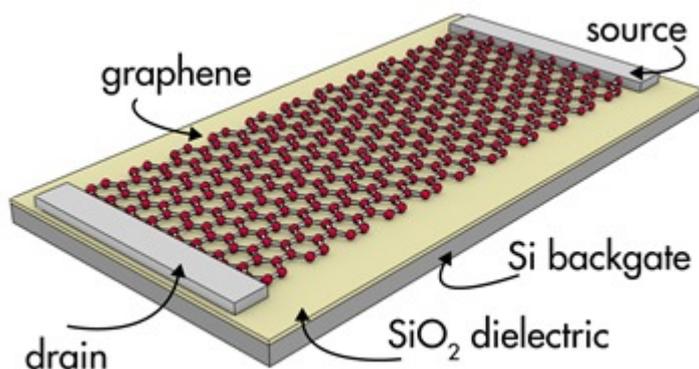
Il est possible d'imiter cette architecture 3D avec des couches métalliques multiples dans les circuits intégrés mais elles coutent pour l'instant assez cher à produire et plus on les empile, plus cela devient compliqué. Les processeurs les plus modernes comprennent une petite dizaine de couches de métallisation, comme indiqué dans ce schéma d'origine Intel.



Il n'est cependant pas théoriquement impossible de superposer des processeurs les uns sur les autres, tout du moins, tant que l'on peut limiter leur réchauffement. L'empilement serait concevable en baissant la fréquence des chipsets, ou avec des techniques extrêmes de refroidissement. Même en divisant par mille la clock des chipsets CMOS, ils resteraient bien plus rapides que la "clock" du cerveau qui est de l'ordre de 100 Hz.

Changer de technologie au niveau des transistors

Cela permettrait d'accélérer leur **vitesse de commutation** et augmenter grâce à cela la fréquence d'horloge des processeurs. Cela peut passer par exemple par des portes au graphène IBM avait **annoncé** en 2011 avoir produit des transistors au graphène capables d'atteindre une fréquence de 155 GHz, et en 40 nm. Les laboratoires qui planchent sur le graphène depuis une dizaine d'année ont bien du mal à le mettre en œuvre en contournant ses écueils et à le fabriquer à un coût raisonnable. Il faudra encore patienter un peu de ce côté-là même si cela semble très prometteur et avec des débouchés dans tous les domaines et pas seulement dans l'IA.



Passer de l'électron au photon

C'est la photonique qui exploite des composants à base des matériaux dits "III-V", un sujet que j'avais exploré dans [Comment Alcatel-Lucent augmente les débits d'Internet en 2013](#). Aujourd'hui, la photonique est surtout utilisée dans le multiplexage de données sur les liaisons ultra-haut-débit des opérateurs télécoms, dans des applications très spécifiques, ainsi que sur des bus de données optiques de supercalculateurs.

La startup française **Lighton.io** planche sur la création d'un coprocesseur optique capable de réaliser très rapidement des calculs sur de gros volumes de données et de combinatoires. Le système s'appuie sur la génération de jeux de données aléatoires permettant de tester simultanément plusieurs hypothèses de calcul, à des fins d'optimisation. Les applications visées sont en premier lieu la génomique et l'Internet des objets.

L'un des enjeux se situe dans l'intégration de composants hybrides, ajoutant des briques en photonique au-dessus de composants CMOS plus lents. Intel et quelques autres sont sur le pont.

Une fois que l'on aura des processeurs optiques généralistes, il faudra relancer le processus d'intégration. Il est actuellement situé aux alentours de 200 nm pour la photonique et la course se déclenche alors pour descendre vers 10 à 5 nm comme pour le CMOS actuel.

Plancher sur les ordinateurs quantiques

Imaginé par le physicien Richard Feynman en 1982, les ordinateurs quantiques sont à même de résoudre certaines classes de problèmes complexes d'optimisation où plusieurs combinatoires peuvent être testées simultanément. Les algorithmes peuvent être résolus de manière polynomiale et non exponentielle. Cela veut dire qu'au gré de l'augmentation de leur complexité, le temps de calcul augmente de manière linéaire avec cette complexité et pas de manière exponentielle. Donc... c'est beaucoup plus rapide !

Mais sauf à être un spécialiste du secteur, on n'y comprend plus rien ! Le principe des qubits qui sous-tendent les ordinateurs quantiques est décrit dans [Quantum computation, quantum theory and AI](#) de Mingsheng Ying, qui date de 2009. Vous êtes très fort si vous comprenez quelque chose à partir de la fin de la seconde page ! Et la presse généraliste et même scientifique simplifie tellement le propos que l'on croit avoir compris alors que l'on n'a rien compris du tout !

Dans [Quantum POMPDs](#), Jennifer Barry, Daniel Barry et Scott Aaronson, du MIT, évoquent en 2014 comment les ordinateurs quantiques permettent de résoudre des problèmes avec des **processus de décision markovien partiellement observables**. Il s'agit de méthodes permettant d'identifier des états optimaux d'un système pour lequel on ne dispose que d'informations partielles sur son état.

Quant à [Quantum Speedup for Active Learning Agents](#), publié en 2014, un groupe de scientifiques espagnols et autrichiens y expliquent comment les ordinateurs quantiques pourraient servir à créer des agents

intelligents dotés de facultés d'auto-apprentissage rapide. Cela serait un chemin vers le développement de systèmes d'IA créatifs.

En 2014, des chinois de l'Université de Sciences et Technologies de Hefei ont été parmi les premiers à **expérimenter** des ordinateurs quantiques pour mettre en jeu des réseaux de neurones artificiels, pour la reconnaissance d'écriture manuscrite. Leur ordinateur quantique utilise un composé organique liquide associant carbone et fluor. On n'en sait pas beaucoup plus !

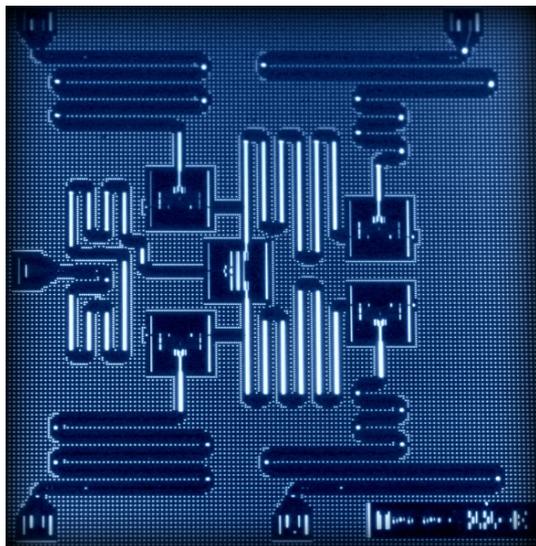
Les équipes de la NASA ont créé de leur côté le QuAIL, le **Quantum Artificial Intelligence Laboratory**, en partenariat avec Google Research. Il utilise un D-Wave Two comme outil d'expérimentation, à ce jour le seul ordinateur quantique commercial, diffusé à quelques unités seulement. Leurs **publications scientifiques** sont abondantes mais pas faciles d'abord comme les autres ! Ce centre de la NASA est situé au Ames Research Center, là-même où se trouve la Singularity University et à quelques kilomètres du siège de Google à Mountain View.

Google annonçait fin 2015 avoir réussi à réaliser des calculs quantiques 100 millions de fois plus rapidement qu'avec des ordinateurs classiques sur ce DWave-Two. Ces tests sont mal documentés au niveau des entrées, des sorties et des algorithmes testés. Il se pourrait même que ces algorithmes soient codés "en dur" dans les qubits des D-Wave ! Qui plus est, la comparaison faite par Google avec les calculs sur ordinateurs traditionnels s'appliquait à algorithme identique alors que les algorithmes utilisés dans l'ordinateur quantique n'étaient pas optimisés pour ordinateurs traditionnels. Bref, le sujet est polémique, comme le rapportent **La Tribune** ou **Science et Avenir**. Est-ce une querelle entre anciens et modernes ? Pas vraiment car ceux qui doutent des performances du D-Wave travaillent aussi sur les ordinateurs quantiques.



“ OUR
QUANTUM COMPUTER
IS
100 MILLION TIMES FASTER
THAN PC.
- GOOGLE

Début mai 2016, **IBM** annonçait mettre à disposition son ordinateur quantique expérimental cryogénique de 5 Qubits en ligne dans son offre de cloud. On ne sait pas trop quel type de recherche pourra être menée avec ce genre d'ordinateur ni quelles APIs sont utilisées.



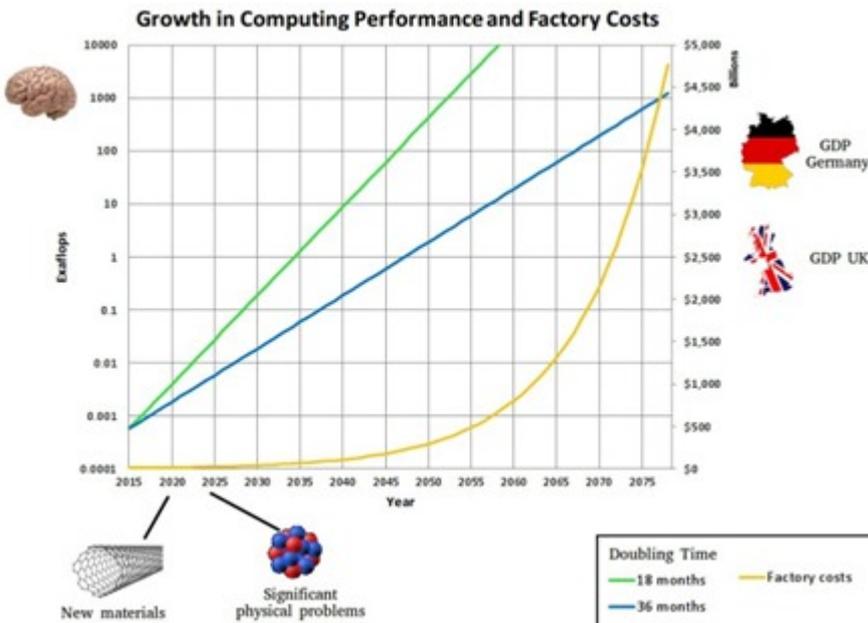
Quid des recherches en France ? Le **CEA** de Saclay planche depuis longtemps sur la création de circuits quantiques. Ils ont développé en 2009 un **dispositif de lecture d'état quantique** non destructif de qubits après avoir créé l'un des premiers qubits en 2002. Et le CEA-LETI de Grenoble a de son côté récemment réalisé des **qubits sur composants CMOS** grâce à la technologie SOI d'isolation des transistors sur le substrat silicium des composants. Ces composants ont toutefois besoin d'être refroidis près du zéro absolu (-273°C) pour fonctionner. Enfin, le groupe français **ATOS**, déjà positionné dans le marché des supercalculateurs depuis son rachat de Bull, travaille avec le CEA pour créer un ordinateur quantique à l'**horizon 2030**.

Dans son étude **Quantum Computing Market Forecast 2017-2022**, le cabinet Market Research Media prévoit que le marché des ordinateurs quantiques fera \$5B d'ici 2020, en intégrant toute la chaîne de valeur matérielle et logicielle. Le premier marché serait celui de la cryptographie. Avant de parler de marché, il faudrait que cela marche ! Et nous n'y sommes pas encore. Chaque chose en son temps : la recherche, l'expérimentation puis l'industrialisation. Nous n'en sommes qu'aux deux premières étapes pour l'instant.

Explorer les ordinateurs moléculaires

Ils permettraient de descendre le niveau d'intégration au-dessous du nanomètre en faisant réaliser les calculs par des molécules organiques de la taille de l'ADN. Cela reste aussi un **animal de laboratoire** pour l'instant ! Mais un animal très prometteur, surtout si l'architecture correspondante pouvait fonctionner de manière tridimensionnelle et plus rapidement que notre cerveau. Reste aussi à comprendre quelle est la vitesse de commutation de ces composants organiques et comment ils sont alimentés en énergie.

Toutes ces innovations technologiques devront surtout se diffuser à un coût raisonnable. En effet, si on extrapole la structure de coût actuelle des superordinateurs, il se pourrait qu'un supercalculateur doté de la puissance du cerveau à une échéance pluri-décennale soit d'un coût supérieur au PIB de l'Allemagne (**source**). Ca calme ! La puissance brute est une chose, son rapport qualité/prix en est une autre !



La notion d'IA intégrative pourrait aussi voir le jour dans les architectures matérielles. Comme le cerveau qui comprend diverses parties spécialisées, un ordinateur doué d'IA évoluée intégrera peut-être des architectures hybrides avec processeurs au graphène, optiques et quantiques en compléments d'une logique de base en bon et vieux CMOS ! Ceci est d'autant plus plausible que certaines techniques sont insuffisantes pour créer un ordinateur générique, notamment les ordinateurs quantiques qui ne sauraient gérer qu'une certaine classe de problèmes, mais pas compresser ou décompresser une vidéo par exemple, ou faire tourner une base de données NoSQL.

Stockage

Si la loi de Moore a tendance à se calmer du côté des processeurs CMOS, elle continue de s'appliquer au stockage. Elle s'est appliquée de manière plutôt stable aux disques durs jusqu'à présent. Le premier disque de 1 To (Hitachi en 3,5 pouces) est apparu en 2009 et on en est maintenant à 8 To. Donc, 2 puissance 4 et Moore est sauf. L'évolution s'est ensuite déplacée vers les disques SSD à mémoires NAND dont la capacité, démarrée plus bas que celle des disques durs, augmente régulièrement tout comme sa vitesse d'accès et le tout avec une baisse régulière des prix. Les perspectives de croissance sont ici plus optimistes qu'avec les processeurs CMOS.

BiCS 3D-NAND

Flash Memory Summit

BiCS delivers smallest chip area of any published 3D-NAND

BiCS U-shaped NAND string enables maximum array efficiency

- Leverages existing NAND Fab infrastructure. Does not need EUV.
- Scaling achieved by increasing number of layers

Good progress in BiCS development

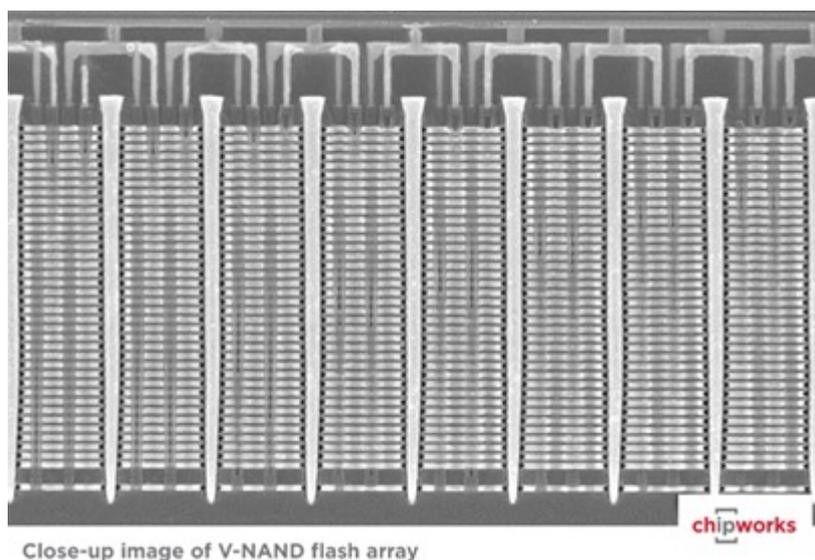
Challenges for all 3D-NAND manufacturing

- NAND poly TFT devices, a first in volume manufacturing
- High aspect ratio etching of large number of layers and its control
- High volume manufacturing requires new etching equipment and techniques for scaling to high number of layers

Note: Diagram not to scale

Comme nous l'avons survolé dans le dernier **Rapport du CES 2016**, les mémoires NAND 3D font des progrès énormes, notamment avec la technologie 3D XPoint d'Intel et Micron qui combine le stockage longue durée et une vitesse d'accès équivalente à celle la mémoire RAM associée aux processeurs. Elle est encore à l'état de prototype mais sa fabrication ne semble pas hors de portée.

La technologie de mémoire 3D est aussi maîtrisée par des sociétés telles que **Samsung** (ci-dessous, avec sa technologique V-NAND) et **Toshiba** (ci-dessus avec sa technologie BiCS). Elle consiste à créer des puces avec plusieurs couches empilées de transistors, ou de transistors montés en colonnes. L'e niveau d'intégration le plus bas des transistors est ici équivalent à celui des CPU les plus denses : il descend jusqu'à 10 nm.



On sait empiler aujourd'hui jusqu'à 48 couches de transistors, et cela pourrait rapidement atteindre une centaine de couches. Des disques SSD de 16 To devraient arriver d'ici peu ! Pourquoi cette intégration verticale est-elle possible pour la mémoire et pas pour les processeurs (GPU, CPU) ? C'est lié à la résistance à la montée en température. Dans un processeur, une bonne part des transistors fonctionne en même temps alors que l'accès à la mémoire est séquentiel et donc n'active pas simultanément les transistors. Un processeur chauffe donc plus qu'une mémoire. Si on empilait plusieurs couches de transistors dans un processeur, il se mettrait à chauffer bien trop et s'endommagerait. Par contre, on sait assembler des circuits les uns sur les autres pour répondre aux besoins d'applications spécifiques.

Pour les supercalculateurs, une tâche ardue est à accomplir : accélérer la vitesse de transfert des données du stockage vers les processeurs au gré de l'augmentation de la performance de ces derniers. Cela va aller jusqu'à intégrer de la connectique à 100 Gbits/s dans les processeurs. Mais la mémoire ne suit pas forcément. Aujourd'hui, un SSD connecté en PCI et avec un connecteur M.2 est capable de lire les données à la vitesse vertigineuse de 1,6 Go/s, soit un dixième de ce qui est recherché dans les calculateurs haute performance (HPC). Mais cette vitesse semble supérieure à celle de lecture d'un SSD ! Le bus de communication est devenu plus rapide que le stockage !

Avec 3D XPoint, l'accès aux données serait 1000 fois plus rapide qu'avec les SSD actuels, modulo l'interface utilisée. Après un retard à l'allumage, cette technologie pourrait voir le jour commercialement en 2017. Elle aura un impact important pour les systèmes d'IA temps réel comme IBM Watson. Rappelons-nous que pour Jeopardy, l'ensemble de la base de connaissance était chargée en mémoire RAM pour permettre un traitement rapide des questions !

Cette augmentation de la rapidité d'accès à la mémoire, qu'elle soit vive ou de longue durée, est indispensable

pour suivre les évolutions à venir de la puissance des processeurs avec l'un des techniques que nous avons examinées juste avant.

Old Constraints

- **Peak clock frequency** as primary limiter for performance improvement
- **Cost:** FLOPs are biggest cost for system: optimize for compute
- **Concurrency:** Modest growth of parallelism by adding nodes
- **Memory scaling:** maintain byte per flop capacity and bandwidth
- **Locality:** MPI+X model (uniform costs within node & between nodes)
- **Uniformity:** Assume uniform system performance
- **Reliability:** It's the hardware's problem

New Constraints

- **Power** is primary design constraint for future HPC system design
- **Cost:** Data movement dominates: optimize to minimize data movement
- **Concurrency:** Exponential growth of parallelism within chips
- **Memory Scaling:** Compute growing 2x faster than capacity or bandwidth
- **Locality:** must reason about data locality and possibly topology
- **Heterogeneity:** Architectural and performance non-uniformity increase
- **Reliability:** Cannot count on hardware protection alone

(source du slide ci-dessus)

Des chercheurs d'université et même de chez Microsoft cherchent à **stocker l'information dans de l'ADN**. Les premières expériences menées depuis quelques années sont prometteuses. La densité d'un tel stockage serait énorme. Son avantage est sa durabilité, estimée à des dizaines de milliers d'années, voire plus selon les techniques de préservation. Reste à trouver le moyen d'écrire et de lire dans de l'ADN à une vitesse raisonnable.

Aujourd'hui, on sait imprimer des bases d'ADN à une vitesse incommensurablement lente par rapport aux besoins des ordinateurs. Cela se chiffre en centaines de bases par heure au grand maximum. Cette vitesse s'accélèrera sans doute dans les années à venir. Mais, comme c'est de la chimie, elle sera probablement plus lente que les changements de phase ou de magnétisme qui ont cours dans les systèmes de stockage numérique actuels. La loi de Moore patientera donc quelques décennies de ce côté là, tout du moins pour ses applications dans le cadre de l'IA.

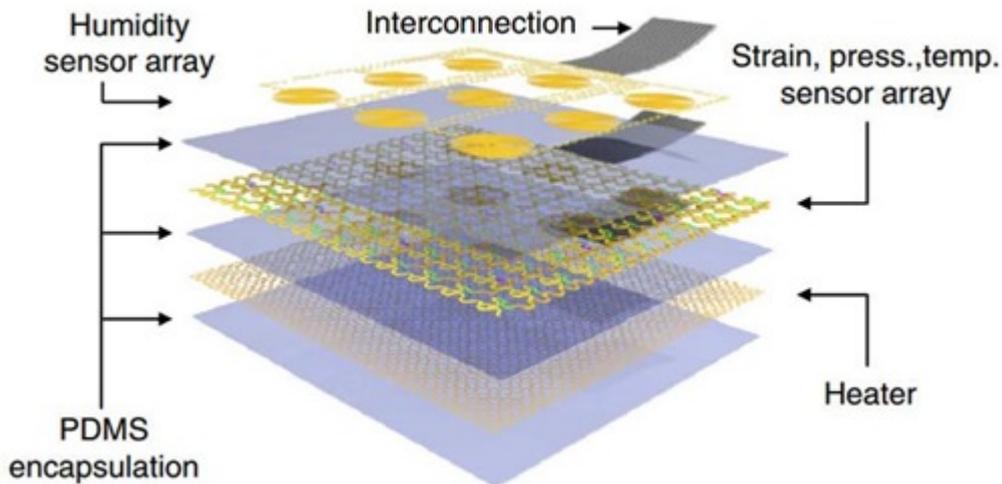
Capteurs sensoriels

L'un des moyens de se rapprocher et même de dépasser l'homme est de multiplier les capteurs sensoriels. La principale différence entre l'homme et la machine réside dans la portée de ces capteurs. Pour l'homme, la portée est immédiate et ne concerne que ses alentours. Pour les machines, elle peut-être distante et globale. On voit autour de soi, on sent la température, on peut toucher, etc. Les machines peuvent capter des données environnementales à très grande échelle. C'est l'avantage des réseaux d'objets connectés à grande échelle, comme dans les "smart cities". Et les volumes de données générés par les objets connectés sont de plus en plus importants, créant à la fois un défi technologique et une opportunité pour leur exploitation.

Le cerveau a une caractéristique méconnue : il ne comprend pas de cellules sensorielles. Cela explique pourquoi on peut faire de la chirurgie à cerveau ouvert sur quelqu'un d'éveillé. La douleur n'est perceptible qu'à la périphérie du cerveau. D'ailleurs, lorsque l'on a une migraine, c'est en général lié à une douleur périphérique au cerveau, qui ne provient pas de l'intérieur. L'ordinateur est dans le même cas : il n'a pas de

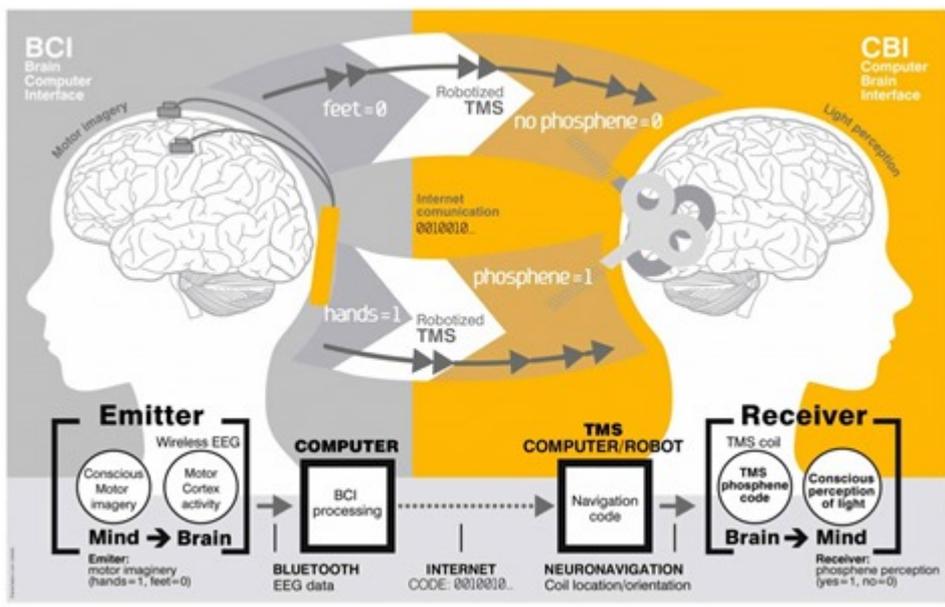
capteurs sensoriels en propre. Il ne ressent rien s'il n'est pas connecté à l'extérieur.

Cette différence peut se faire sentir même à une échelle limitée comme dans le cas des véhicules à conduite assistée ou automatique qui reposent sur une myriade de capteurs : ultrasons, infrarouges, vidéo et laser / LIDAR, le tout fonctionnant à 360°. Ces capteurs fournissent aux ordinateurs de bord une information exploitable qui va au-delà de ce que le conducteur peut percevoir. C'est l'une des raisons pour lesquelles les véhicules automatiques sont à terme très prometteurs et plus sécurisés. Ces techniques sont déjà meilleures que les sens humains, surtout en termes de temps de réponse, de vision à 360° et de capacité d'anticipation des mouvements sur la chaussée (piétons, vélos, autres véhicules).

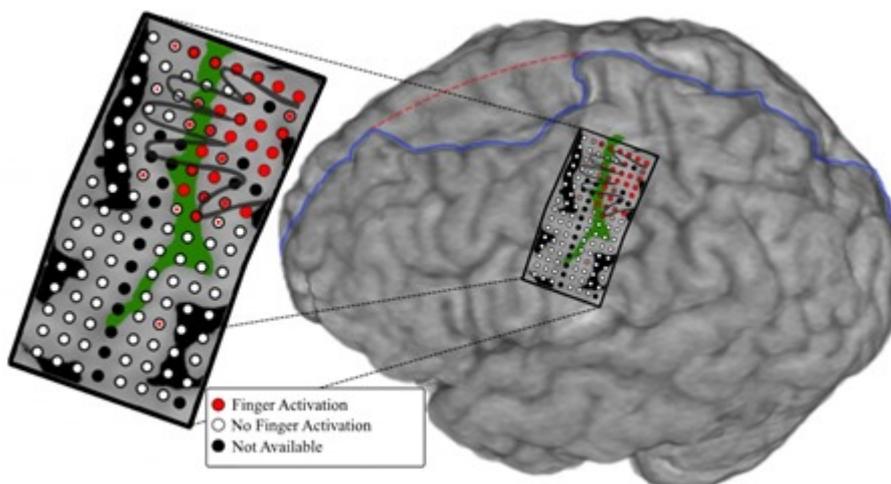


Les capteurs de proximité intégrables à des machines comme les robots progressent même dans leur biomimétisme. Des prototypes de peau artificielle sensible existent déjà en laboratoire, comme en Corée du Sud (ci-dessous, **source dans Nature**). L'une des mécaniques humaines les plus difficiles à reproduire sont les muscles. Ils restent une mécanique extraordinaire, économe en énergie, fluide dans le fonctionnement, que les moteurs des robots ont bien du mal à imiter.

Les capteurs fonctionnent aussi dans l'autre sens : de l'homme vers la machine. Les progrès les plus impressionnants concernent les capteurs cérébraux permettant à l'homme de contrôler des machines, comme pour contrôler un membre artificiel robotisé, une application pouvant restaurer des fonctions mécaniques de personnes handicapées, voire de démultiplier la force de personnes valides, dans les applications militaires ou de BTP. L'homme peut ainsi piloter la machine car la périphérie du cortex cérébral contient les zones où nous commandons nos actions musculaires. Des **expériences de télépathie** sont également possibles, en captant par EEG la pensée d'un mot d'une personne et en la transmettant à distance à une autre personne en lui présentant ce mot sous forme de flash visuel par le procédé TMS, de stimulation magnétique transcraniale.



Si on peut déjà alimenter le cerveau au niveau de ses sens, comme de la vue, en interceptant le nerf optique et en simulant le fonctionnement de la rétine ou par la TMS, on ne sait pas l'alimenter en **idées** et **informations abstraites** car on ne sait pas encore vraiment comment et surtout où elles sont stockées. Dans *Mashable*, une certaine Marine Benoit affirmait un peu rapidement en mars 2016 qu'une équipe avait mis au point "*un stimulateur capable d'alimenter directement le cerveau humain en informations*". A ceci près que l'étude en question, *Frontiers in Human Neuroscience* ne faisait état que d'un système qui modulait la capacité d'acquisition par stimulation ! Pour l'instant, on doit se contenter de lire dans le cerveau dans la dimension mécanique mais pas "écrire" dedans directement. On ne peut passer que par les "entrées/sorties", à savoir les nerfs qui véhiculent les sens, mais pas écrire directement dans la mémoire. Mais ce n'est peut-être qu'un début !



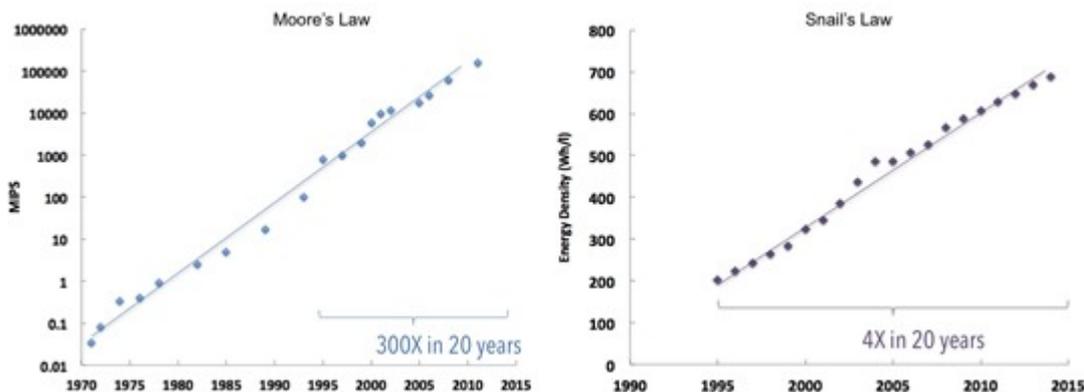
(source de la photo, crédit Guy Hotson)

Energie

L'homme ne consomme en moyenne que 100 Watts dont 20 Watts pour le cerveau. C'est un excellent rendement. Tout du moins, pour ceux qui font travailler leur cerveau. Ce n'est pas facile à égaler avec une machine et pour réaliser les tâches de base que réalise un humain. Les supercalculateurs consomment au mieux quelques KW et certains dépassent les MW.

Des progrès sont cependant notables dans les processeurs mobiles. Consommant moins de 5 W, ils agrègent une puissance de calcul de plus en plus impressionnante grâce à des architectures multi-cœurs, à un fonctionnement en basse tension, aux technologies CMOS les plus récentes comme le FinFET (transistors verticaux) ou FD-SOI (couche d'isolant en dioxyde de silicium réduisant les fuites de courant dans les transistors et améliorant leur rendement énergétique) et à une fréquence d'horloge raisonnable (entre 1 et 1,5 GHz).

La mécanique et l'énergie sont les talons d'Achille non pas de l'IA qui est distribuable là où on le souhaite mais des robots. Un homme a une autonomie d'au moins une journée en état de marche convenable sans s'alimenter. Un robot en est encore loin. D'où l'intérêt des travaux pour améliorer les batteries et notamment leur densité énergétique. Un besoin qui se fait sentir partout, des smartphones et laptops aux véhicules électriques en passant par les robots. Les progrès dans ce domaine ne sont pas du tout exponentiels. Cela a même plutôt tendance à stagner. Dans les batteries, c'est la loi de l'escargot qui s'appliquerait avec un quadruplement de la densité tous les 20 ans (source).



Des laboratoires de recherche inventent régulièrement des technologies de batteries battant des records en densité énergétique ou du côté du temps de chargement, à base de matériaux différents et/ou de nano-matériaux. Mais en elles sortent rarement, faute de pouvoir être industrialisées à un coût raisonnable ou de bien fonctionner dans la durée. Parfois, on arrive à une densité énergétique énorme, mais cela ne fonctionne que pour quelques cycles de charge/décharge. Trop injuste !

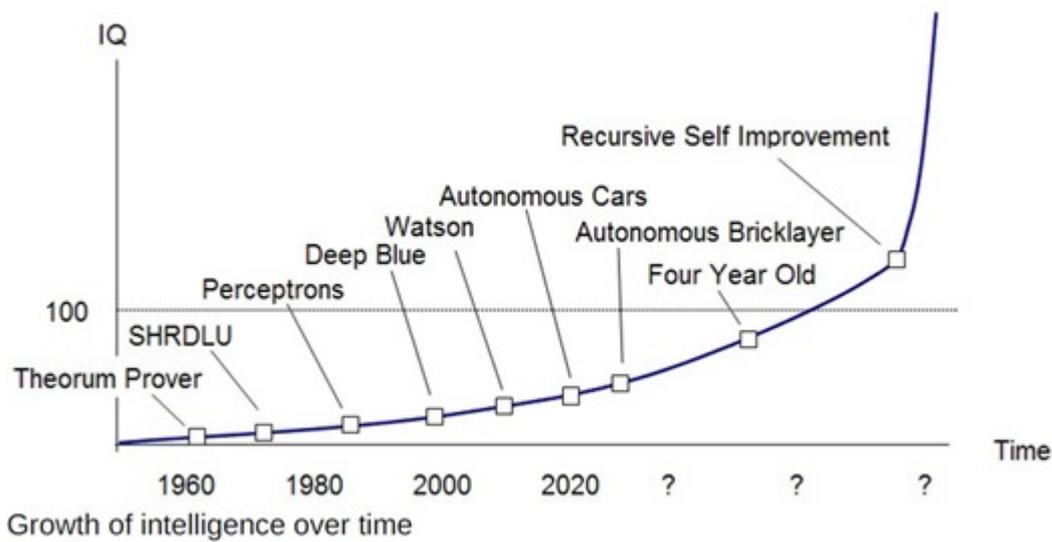
Résultat, pour le moment, la principale voie connue est celle de l'efficacité industrielle, choisie par Elon Musk dans la création de sa Gigafactory dans le Nevada, une usine à \$5B qui exploitera la technologie de batteries standards de Panasonic, qui a aussi mis \$1B au pot pour le financement de l'usine. Une usine qui est aussi proche d'une mine de Lithium, à Clayton Valley, l'un des composés clés des batteries et qui démarrera sa production en 2020.

On peut cependant citer l'étonnante performance d'un laboratoire de l'université de Columbia qui a réussi à alimenter un composant CMOS avec de l'énergie provenant de l'ATP (adénosine triphosphate), la source d'énergie principale des cellules vivantes qui est générée par les nombreuses mitochondries qu'elles contiennent. Cela ouvre des portes vers la création de solutions hybrides biologiques et informatiques insoupçonnées jusqu'à présent.

Sécurité

C'est un sujet évoqué de manière indirecte, au sujet du jour où l'IA dépassera l'intelligence de l'homme et s'auto-multipliera au point de mettre en danger l'espèce humaine. Cela part du principe qu'une intelligence peut se développer à l'infini in-silico. Pourquoi pas, dans certains domaines. Mais c'est faire abstraction d'un point clé : l'intelligence est le fruit, certes, du fonctionnement du cerveau, mais également de l'interaction avec l'environnement et avec les expériences sensorielles. L'intelligence cumule la capacité à créer des théories expliquant le monde et à des expériences permettant de le vérifier. Parfois, la vérification s'étale sur un demi-

siècle à un siècle, comme pour les ondes gravitationnelles ou le Boson de Higgs. Cette capacité de théorisation et d'expérimentation de long terme n'est pour l'instant pas accessible à une machine, quelle qu'elle soit.



(schéma tiré de "The artificial intelligence singularity, 2015")

L'IA présente des risques bien plus prosaïques, comme toutes les technologies numériques : dans sa sécurité. Celle d'un système d'IA peut être compromise à plusieurs niveaux : dans les réseaux et le cloud, dans les capteurs, dans l'alimentation en énergie. Les bases de connaissances peuvent aussi être induites en erreur par l'injection d'informations erronées ou visant à altérer le comportement de l'IA, par exemple dans le cadre d'un diagnostic médical complexe. On peut imaginer l'apparition dans le futur d'anti-virus spécialisés pour les logiciels de machine learning.

Les dangers de l'IA, s'il en existe, sont particulièrement prégnants dans l'interaction entre les machines et le monde extérieur. Un robot n'est pas dangereux s'il tourne en mode virtuel dans une machine. Il peut le devenir s'il tient une arme dans le monde extérieur et qu'il est programmé par des forces maléfiques. Le "kill switch" de l'IA qui permettrait de la déconnecter si elle devenait dangereuse devrait surtout porter sur sa relation avec le monde physique. Les films de science fiction comme Transcendance montrent que rien n'est sûr de ce côté là et que la tendance à tout automatiser peut donner un trop grand contrôle du monde réel aux machines.

L'homme est déjà dépassé par la machine depuis longtemps, d'abord sur la force physique, puis de calcul, puis de mémoire et enfin de traitement. Mais la machine a toujours été pilotée par l'homme. L'IA semble générer des systèmes pérennes dans le temps *ad vitam aeternam* du fait de processus d'apprentissage qui s'agrègent avec le temps et de la mémoire presque infinie des machines. L'IA serait immortelle. Bon, tant que son stockage ne plante pas ! Un disque dur peut planter à tout bout de champ au bout de cinq ans et un disque SSD actuel ne supporte au mieux que 3000 cycles d'écriture !



Les dangers perceptibles de l'IA sont à l'origine de la création d'OpenAI, une initiative visant non pas à créer une IA open source (cela existe déjà dans le machine learning) mais de surveiller ses évolutions. Il s'agit d'une ONG créée par Elon Musk qui vise à s'assurer que l'IA fasse le bien et pas le mal à l'humanité. Elle est dotée de \$1B et doit faire de la recherche. Un peu comme si une organisation était lancée pour rendre le capitalisme responsable (cf **OpenAI** dans Wikipedia et “Why you should fear artificial intelligence” paru dans TechCrunch en mars 2016).

Autre méthode, se rassurer avec “**Demystifying Machine Intelligence**” de Piero Scaruffi qui cherche à démontrer que la singularité n'est pas pour demain. Il s'appuie pour cela sur une vision historique critique des évolutions de l'intelligence artificielle. Il pense que les progrès de l'IA proviennent surtout de l'augmentation de la puissance des machines, et bien peu des algorithmes, l'effet donc de la force brute. Selon lui, l'homme a toujours cherché une source d'intelligence supérieure, qu'il s'agisse de dieux, de saints ou d'extra-terrestres. La singularité et les fantasmes autour de l'IA seraient une nouvelle forme de croyance voire même de religion, une thèse aussi partagée par Jaron Lanier, un auteur anticonformiste qui publiait “**Singularity is a religion just for digital geeks**” en 2010.

Singularity Is a Religion Just for Digital Geeks

by JARON LANIER

Piero Scaruffi prend aussi la singularité à l'envers en avançant que l'ordinateur pourra dépasser l'homme côté intelligence parce que les technologies rendent l'homme plus bête, en le déchargeant de plus en plus de fonctions intellectuelles, la mémoire en premier et le raisonnement en second ! Selon lui, le fait que les médias numériques entraînent les jeunes à lire de moins en moins de textes longs réduirait leur capacité à raisonner. On peut d'ailleurs le constater dans les débats politiques qui évitent la pensée complexe et privilégient les simplismes à outrance. J'aime bien cet adage selon lequel l'intelligence artificielle se définit comme étant le contraire de la bêtise naturelle. Cette dernière est souvent confondante et rend le défi de la création d'une intelligence artificielle pas si insurmontable que cela.

Pour Piero Scaruffi, en tout cas, l'intelligence artificielle est d'ailleurs une mauvaise expression. Il préfère évoquer la notion d'intelligence non humaine. Il pense aussi qu'une autre forme d'intelligence artificielle pourrait émerger : celle d'hommes dont on aura modifié l'ADN pour rendre leur cerveau plus efficace. C'est un projet du monde réel, poursuivi par les chinois qui séquençent des milliers d'ADN humains pour identifier les

gènes de l'intelligence ! Histoire de réaliser une (toute petite) partie des fantasmes délirants du film Lucy de Luc Besson !

Pour **Daniel C. Dennett**, le véritable danger n'est pas dans les machines plus intelligentes que l'homme mais plutôt dans le laisser-aller de ce dernier qui abandonne son libre arbitre et confie trop de compétences et d'autorité à des machines qui ne lui sont pas supérieures.

Et si le plus grand risque était de ne rien faire ? Pour toutes ces technologies et recherches citées dans cet article, est-ce que l'Europe et la France jouent un rôle moteur ? Une bonne part de cette R&D côté hardware est concentrée au CEA. Pour l'industrie, ce n'est pas évident, à part peut-être la R&D en photonique chez Alcatel-Lucent qui même si elle dépend maintenant de Nokia, n'en reste pas moins toujours en France. Il reste aussi STMicroelectronics qui reste très actif dans les capteurs d'objets connectés. De son côté, la R&D côté logicielle est dense, que ce soit à l'INRIA ou au CNRS. Reste à savoir quelle "technologie de rupture" sortira de tout cela, et avec une transformation en succès industriel à grande échelle qui passe par de l'investissement, de l'entrepreneuriat et de la prise de risque car de nombreux paris doivent être lancés en parallèle pour n'en réussir que quelques-uns.

Dans la **partie suivante**, la dernière de cette série, nous allons nous intéresser à la robotisation des métiers, en prenant un peu de recul par rapport à la vulgate actuelle qui prédit la disparition de la moitié des métiers d'ici deux décennies !

Vous pouvez consulter tous les épisodes de ce roman fleuve de printemps sur l'intelligence artificielle :

Episode 1 : **sémantique et questions clés**

Episode 2 : **histoire et technologies de l'intelligence artificielle**

Episode 3 : **IBM Watson et le marketing de l'intelligence artificielle**

Episode 4 : **les startups US de l'intelligence artificielle**

Episode 5 : **les startups acquises par les grands du numérique**

Episode 6 : **les startups françaises de l'intelligence artificielle**

Episode 7 : **la modélisation et la copie du cerveau**

Episode 8 : **évolutions de la loi de Moore et applications à l'intelligence artificielle**

Episode 9 : **la robotisation en marche des métiers**

Cet article a été publié le 2 mai 2016 et édité en PDF le 23 mars 2024.
(cc) Olivier Ezratty – "Opinions Libres" – <https://www.oezratty.net>