



## Peut-on benchmarker l'intelligence artificielle ?

L'intelligence artificielle est l'un des grands sujets technologiques à la mode depuis 2016. L'année avait démarré sur les chapeaux de roue avec la victoire de Google DeepMind AlphaGo sur le champion mondial du jeu de Go. Les investissements dans les startups de l'IA ont connu un pic en 2015 et 2016. L'intelligence artificielle a même remplacé la mobilité dans les priorités des DSI, selon les grands cabinets d'analystes. On est passé du "mobile first" à "l'IA first". Les analystes envoient un message d'urgence vers les grandes entreprises, les enjoignant d'intégrer fissa l'IA dans leur stratégie, sous peine de se voir disrupter par d'autres et de devenir obsolètes. Ce n'est plus de l'uberisation, symbole de nouvelles formes d'intermédiation, mais le remplacement par des services qui ont moins besoin de salariés pour fonctionner.

Cela fait maintenant plus d'un an que je potasse le sujet. Entre mars et mai 2016, j'avais publié une série de 9 articles de défrichage de l'IA, compilés ensuite sous la **forme d'un ebook**. Depuis 2015, un nombre incalculable d'ebooks et livres blancs ont été publiés par divers analystes et éditeurs de solutions.

S'y est ajoutée l'imposante compilation du plan **France IA**, pilotée par Bercy et le Ministère de la Recherche. C'est un excellent tour d'horizon de l'existant en France tant côté recherche que de celui des entreprises. J'avais il y a deux mois fait un tour en plein et en creux de ce plan. Il se trouve que ce n'était qu'un plan d'étape. Mais les germes d'une dispersion restent toujours présents dans les approches qui consolident les avis de centaines de parties prenantes. C'est à l'image du sujet qui est en fait aussi large que le numérique ou le logiciel.

L'IA est aussi devenu un objet éminemment politique, soulevant de nombreuses questions d'éthique et aussi, sur les évolutions du marché de l'emploi qu'elle va entraîner. C'est notamment l'objet du rapport parlementaire **Pour une intelligence artificielle maîtrisée, utile et démystifiée** finalisé et publié à peu près au même moment que le plan France IA, en mars 2017.

J'ai eu vent récemment de la demande de quelques grands comptes français demandant à des sociétés de conseil de réaliser un "benchmark de l'IA". Un peu comme si c'était un produit en étagères, comme un banal serveur de base de données. Les bras m'en sont un peu tombés. La demande n'était pas pour autant aussi farfelue que cela : il est bon de se demander ce que l'IA pourrait apporter à une entreprise et d'en évaluer le potentiel. Reste à déterminer la bonne méthode.



Il se trouve que je me suis encore plus plongé dans les arcanes scientifiques et technologiques de l'IA, aussi bien du côté du deep learning que de l'IA symbolique ou des avancées dans le domaine de l'intelligence artificielle générale (AGI). Je me suis rendu compte à quel point le champ de recherche et d'applications de l'IA étaient immenses et aussi, totalement incompris par une bonne part des entreprises et de l'écosystème. Je ne suis d'ailleurs pas encore sûr d'avoir tout compris ! L'incomplétude de la compréhension est permanente dans ce domaine.

Il n'est en tout cas pas du tout monolithique. Si l'on ne prend que le le deep learning s'appuyant sur des réseaux de neurones, cela avance régulièrement et des chercheurs font évoluer chaque année le sujet. L'IA préserve encore un côté mystérieux et magique car ses méthodes mathématiques et logiques ne sont pas du plus facile des abords. Qui plus est, les prévisions de l'augmentation de la puissance de l'IA s'appuient encore de manière un peu simpliste sur une application sans limites de la loi de Moore.

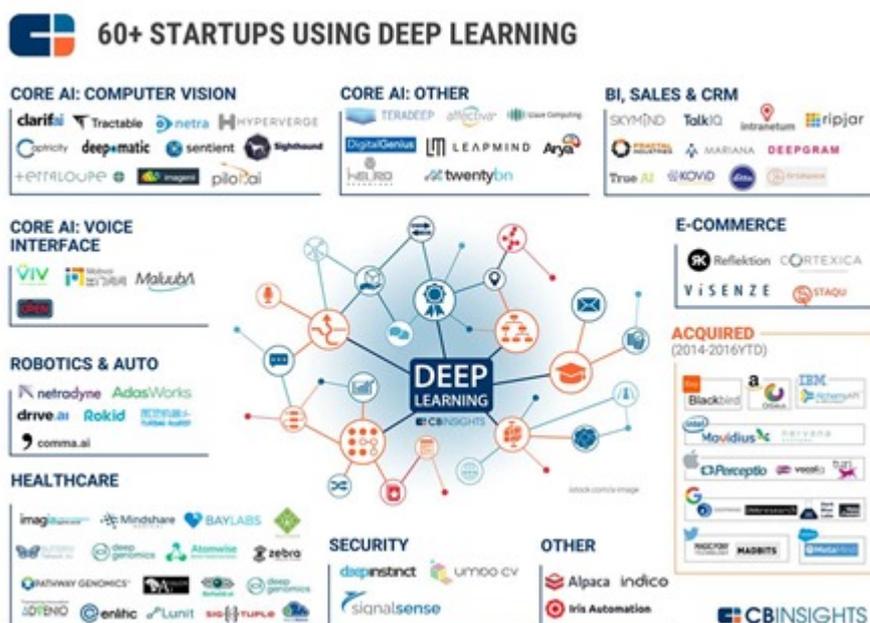
## L'outillage en IA

Pour comprendre la structure d'un projet d'IA, on peut commencer par observer de près ce que font couramment les startups de l'intelligence artificielle. Dans la pratique, elles ne créent pas beaucoup de logiciels. Elles sont surtout des assembleurs de données dans un domaine précis – santé, finance, marketing, juridique, autre – qu'elles injectent dans des moteurs de machine learning ou de deep learning du marché.

La qualité de leurs solutions provient surtout de celle des données qui les alimentent. Parfois, l'IA de ces startups prend même la forme d'une boîte noire, notamment pour les applications de reconnaissance d'images, utilisées en particulier pour la réalisation de diagnostics automatisés (fonds de l'œil et dermatologie dans l'imagerie médicale, contrôle qualité en usine pour l'industrie manufacturière). Ces boîtes noires sont fournies par de grandes startups américaines telles que **Clarifai.ai**.

En pratique, que pourrait signifier "benchmarker l'IA" dans une entreprise ? On peut benchmarker d'une part des outils et, d'autre part, leur mise en œuvre dans des projets précis utilisant des jeux de données de tests. Certains outils peuvent être prêts à l'emploi, comme la ribambelle de chatbots et autres robots conversationnels grand publics qui ont pullulé en 2015 et 2016 aussi bien aux USA qu'en France.

Vous pouvez obtenir de nombreux panoramas de startups du secteur grâce aux excellentes cartographies de l'américain **CBInsight**. Elles couvrent aussi bien les startups de technologies génériques que de leurs applications dans des marchés verticaux et horizontaux.



Pour les entreprises qui souhaitent s'outiller elles-mêmes, il existe un grand nombre de solutions open source pour lancer un projet d'IA, et notamment dans le deep learning. L'un d'entre eux se détache du lot et est très utilisé par les startups : la bibliothèque open source **TensorFlow** créée par Google, qui fonctionne aussi bien en embarqué que dans le cloud. Elle est mise en œuvre par quasiment les trois quarts des startups françaises de

l'IA que j'ai pu croiser ces derniers mois. Derrière TensorFlow, l'un des plus courus est **Theano**, issu de l'Université de Montréal.

## principaux frameworks de deep learning



Deux outils créés par des français sont aussi couramment utilisés : **Scikit-Learn** et **Keras**, ce dernier étant un sur-ensemble de TensorFlow et Theano. Ces frameworks sont généralement utilisés via du code écrit en **Python**. Certains de ces frameworks sont comparés dans cette intéressante présentation : **Deep Learning Frameworks : TensorFlow, Theano, Keras, Torch, Caffe** datant de janvier 2017. Pour les courageux, la **documentation de Scikit-Learn** fait 2125 pages ! On ne rentre pas là-dedans comme pour apprendre à programmer en HTML/CSS et même PHP ! Un projet standard va généralement nécessiter de choisir un grand nombre d'outils dans une chaîne de valeur complexe.

	Languages	Tutorials and training materials	CNN modeling capability	RNN modeling capability	Architecture: easy-to-use and modular front end	Speed	Multiple GPU support	Keras compatible
Theano	Python, C++	++	++	++	+	++	+	+
TensorFlow	Python	+++	+++	++	+++	++	++	+
Torch	Lua, Python (new)	+	+++	++	++	+++	++	
Caffe	C++	+	++		+	+	+	
MXNet	R, Python, Julia, Scala	++	++	+	++	++	+++	
Neon	Python	+	++	+	+	++	+	
CNTK	C++	+	+	+++	+	++	+	

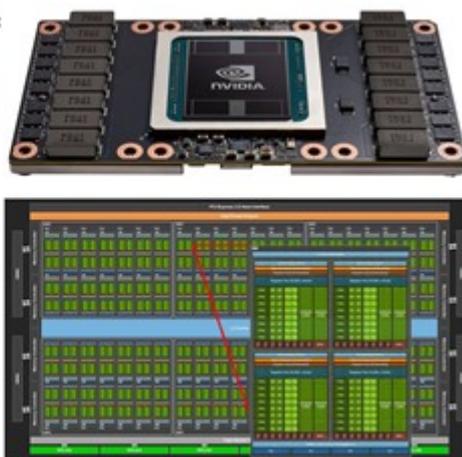
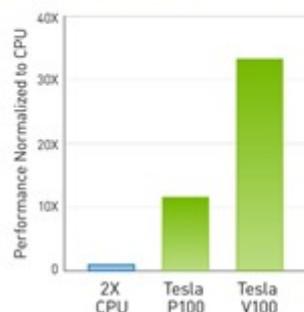
**TensorFlow** ressort souvent du lot car c'est l'un des frameworks qui bénéficie du plus grand nombre de contributions sur Github et qu'il bénéficie du soutien de Google. Le schéma ci-dessus met en avant cette différence (source). Il requiert un petit décodage : les **CNN** sont les réseaux de neurones à convolution (qui découpent un problème en sous-problèmes pour analyser par exemple le contenu d'une image en la décomposant par éléments ayant un niveau d'abstraction croissant), les **RNN** (réseaux de neurones récurrents, utilisés dans la reconnaissance de la parole, et dans la production de prédictions). Tous ces moteurs de deep learning ont des capacités variées parmi lesquelles il faudra choisir son ou ses modèles de réseaux de neurones en fonction de la nature du besoin !



TensorFlow est aussi capable de bien monter en puissance pour les gros volumes de données et d'exploiter des architectures en cloud reposant sur des GPU comme ceux de **Nvidia**. Ce dernier vient d'ailleurs tout juste d'annoncer une nouvelle génération de GPU dédié au deep learning, le GV100. Produit en technologie 12 nm chez TSMC, il intègre la bagatelle de 21,1 milliards de transistors dans 5376 cœurs, le tout ayant une puissance de calcul impressionnante de 120 TFLOPS.

**Nvidia Tesla V100 avec processeur GV100 :**  
21,1 milliards de transistors, 12 nm, 120 TFLOPS, 5376 cœurs, 815 mm<sup>2</sup>

30x Higher Throughput than CPU  
Server on Deep Learning Inference



Mais l'IA ne se résume pas au deep learning. Celui-ci est souvent mis en œuvre dans les applications de vision artificielle pour tagger automatiquement des photos ou des vidéos ou pour identifier des pathologies dans de l'imagerie médicale. On peut utiliser des techniques de machine learning et de deep learning pour segmenter automatiquement une base de données de clients, ce que l'on appelle la "clusterisation". Ils servent aussi à faire des prédictions s'appuyant sur des jeux de données complexes. On les a vus à l'œuvre pour les prédictions électorales s'appuyant par exemple sur les conversations dans les réseaux sociaux. Avec plus ou moins de bonheur d'ailleurs selon que l'électorat jeune ou moins jeune déterminait l'issue d'une élection.

Il existe aussi des moteurs d'**IA symbolique**, assimilables aux systèmes experts des années 1980. On les alimente avec des corpus de règles pour ensuite, les faire répondre à des questions et traiter des problèmes complexes nécessitant une logique inductive ou déductive. Cette forme d'IA s'appuie sur des connaissances structurées. La société **Cycorp** propose ainsi un corpus de 630 000 concepts, 7 000 000 faits et règles et 38 000 relations couvrant de larges pans du savoir humain générique. Il a été accumulé sur plusieurs décennies. Les règles peuvent être aussi extraites de données non structurées comme des bases textuelles (genre Wikipedia) ou les réseaux sociaux. Quelques startups proposent des outils dans ce domaine, comme **Vicarious** ou **Numenta**. Il existe aussi des briques associées dans des architectures plus vastes telles que **IBM Watson**.

Bref, les frameworks d'IA et les infrastructures techniques qui les exécutent sont les seules technologies que l'on peut benchmarker de manière traditionnelle, en comparant entre elles des solutions techniques et à divers niveaux : langage, framework, moteur, outils de préparation des données, ressources en cloud, outils de parallélisation des traitements (**Hadoop**, Amazon **Spark**), serveurs et processeurs (CPU, GPU, processeurs synaptiques mettant directement en œuvre des réseaux de neurones).

## Les données

Dans le contexte d'un projet d'entreprise, un projet d'IA démarre souvent avec des données et si possible avec de gros volumes de données. Le volume et la qualité des données sont clés pour bien entraîner un moteur de deep learning. C'est l'une des raisons de la force des GAFAs : ils ont naturellement accès à d'immenses volumes

de données liées aux actions des utilisateurs de Google Search, Facebook, iOS, Android, SIRI, Amazon Alexa, etc. Les sociétés qui déploient de gros volumes d'objets connectés ont aussi accès à des données intéressantes à exploiter.

Un benchmark d'entreprise doit donc partir d'un ou de jeux de données dont on veut extraire quelque chose.

Il faut bien évidemment se poser la question de ce que l'on veut en faire. Au départ, on ne sait pas trop. L'entreprise dispose par exemple d'une base de données du comportement de ses clients et voudrait l'utiliser pour identifier les clients à potentiel d'upsell ou de cross-sell (ventes additionnelles), ou au contraire, ceux qui peuvent générer du churn (abandonner l'offre). Elle peut aussi vouloir déterminer les actions à mener pour optimiser un système complexe : client, production, autre. L'IA peut aussi servir dans tout un tas de domaines : dans la robotique (qui intègre généralement tout un tas de briques technologiques : vision artificielle, mécanique, systèmes experts, etc), dans la relation client, pour créer des solutions de recommandation, pour analyser des tendances, pour analyser l'image de l'entreprise dans les médias et les réseaux sociaux. Etc. Et la gradation est forte entre générique et spécifique dans ces différentes solutions.

Des projets d'IA peuvent se passer de machine learning et de deep learning et s'appuyer sur des connaissances structurées et des moteurs de règles. C'est par exemple le cas pour créer des systèmes d'assistance à la maintenance industrielle. Dès lors que l'on manipule des données très structurées et une architecture de concepts, les outils de deep learning sont inadaptés. On se retrouve ici dans un domaine ancien, qui a connu ses heures de gloire pendant les années 1980, avec LISP et Prolog. Il n'est pas périmé pour autant, malgré tout le tintouin autour du deep learning, présenté à tort comme une sorte de solution universelle des besoins de l'IA.

Comme pour toute nouvelle technologie, l'innovation procède de manière itérative en découvrant et... en comprenant, ce que la technologie peut apporter et en identifiant des besoins non pourvus par les technologies traditionnelles qui pourraient être traités par ces nouveaux outils un peu mystérieux. Cette démarche nécessite de faire appel à des spécialistes de l'IA qui connaissent la diversité de ses nombreuses techniques et méthodes.

Conceptuellement, pour les entreprises qui disposent de gros volumes de données, l'IA constitue souvent un ensemble de techniques qui complète une longue lignée de technologies : les bases de données, la business intelligence, le big data, les data analytics et la data intelligence.

## **Le projet**

Supposons qu'un besoin est bien identifié, que les données soient disponibles et que les technologies de l'IA puissent apporter une solution. Une fois ceci qualifié, on peut rentrer en mode projet.

Va-t-on mener plusieurs "proof of concepts" en parallèle avec différents frameworks d'IA tels que ceux qui sont cités auparavant, et comparer ensuite les résultats ? Ce ne serait pas raisonnable et surtout, ce serait trop coûteux. Il vaut mieux se faire conseiller pour choisir les bons outils et ensuite mener son expérimentation.

Un projet d'IA d'entreprise a un petit côté "recherche applicative". En effet, on ne va pas juste alimenter une bête de deep learning avec un tombereau de données et attendre un beau résultat à la sortie d'un tuyau. Il va falloir extraire et préparer les données, les nettoyer, les filtrer, savoir ne conserver que ce qui est pertinent. On va ensuite paramétrer les frameworks de deep learning ou autres en fonction des algorithmes à utiliser. Il n'existe pas une technique unifiée de machine learning ou de deep learning, mais des dizaines de variantes ! Puis on va observer les résultats. Ils ne seront pas forcément probants du premier coup. Il faudra reboucler sur les données et le paramétrage pour affiner le modèle. Et il faudra aussi bien visualiser les résultats pour qu'ils soient compréhensibles.

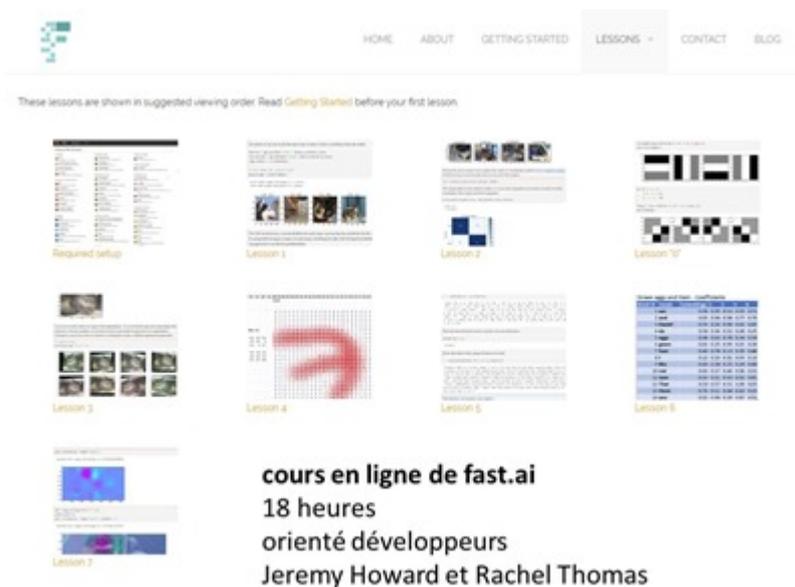
On appréciera alors la qualité des résultats. Le propre des techniques de machine learning et de deep learning

est qu'elles génèrent rarement des résultats exacts à 100%. Il y a toujours un taux d'erreur, que l'on minimise avec l'expérience et que l'on cherche à faire descendre en-dessous d'un niveau acceptable. Comme le taux d'erreurs d'un système de reconnaissance vocale ou celui de l'identification de pathologies dans de l'imagerie médicale. A ce jour, les solutions les plus avancées dans ce dernier domaine génèrent un taux d'erreur plus faible que celui des spécialistes ! C'est donc acceptable !

Un benchmark pourra éventuellement avoir lieu pour comparer un projet mené en mode "IA" et un projet mené avec des outils traditionnels de data mining. Si ceux-ci peuvent donner des résultats sur des données chiffrées, ils ne seront pas du tout à la hauteur pour traiter des données images/vidéo/audio, là où le deep learning est devenu indispensable.

## Les équipes

Un projet d'IA est comme un projet d'objets connectés : il va devoir réunir des talents et compétences divers, certaines internes aux entreprises, d'autres externes. Il y a la compétence métier qui prime. Suit la compétence IT plus traditionnelle, pour la collecte et l'exploitation des données. Le paramétrage des moteurs d'IA passe par des spécialistes d'un nouveau genre qui ont de bonnes bases en IA sachant que la France en forme à peine 1300 par an actuellement. Ils sont complétés par des "data scientists" qui jouent parfois tous les rôles. Après avoir rencontré les pires des difficultés à recruter de bons développeurs, les entreprises de services, les éditeurs de logiciels comme les startups vont rencontrer de grandes difficultés à recruter les bons talents à même de paramétrer un moteur de deep learning ! Cela va nécessiter une mise à niveau de nombreux cursus de formation ! Les plus courageux peuvent se former en ligne comme avec ces **18 heures de cours en ligne** de Fast.ai, réalisées par Jeremy Howard et Rachel Thomas.



These lessons are shown in suggested viewing order. Read [Getting Started](#) before your first lesson.

**cours en ligne de fast.ai**  
18 heures  
orienté développeurs  
Jeremy Howard et Rachel Thomas

Les entreprises de services numériques qui maîtrisent ce genre de projet ne sont pas encore nombreuses. Elles sont en train de s'y mettre. Les startups ne sont de leur côté pas forcément adaptées à la conduite de projets. Sauf pour gagner les premiers clients en entreprise. Pour les repérer, on peut commencer par visiter leurs sites web et inventorier leurs représentants chefs de projets et ingénieurs qui s'expriment dans les conférences sur l'IA. Dans les mois et années à venir, les grands acteurs du service vont probablement faire l'acquisition de petits acteurs spécialisés dans l'IA.

Les projets peuvent être vite couteux s'il faut mettre en branle une armée de consultants, data scientists, développeurs et aussi designers. Même si le cœur du réacteur d'un projet d'IA est spécifique à l'IA, autour, il

faudra aussi faire tourner des briques plus classiques, tant côté back-end (préparation des données, bases de données, stockage, infrastructure) que du front-end (créer de belles interfaces pour les utilisateurs).

### Les études de cas

L'autre forme de benchmark consiste à analyser des études de cas de clients se rapprochant de son secteur d'activité. Les grands acteurs de l'IA en publient quelques-unes.

L'un des plus prolixes dans l'exercice est IBM qui promeut ses solutions et services de "cognitive computing" via la marque IBM Watson. Ces différentes études de cas ne sont cependant pas toujours très bien documentées. La plupart d'entre elles sont situées au niveau marketing, très simpliste avec peu de précisions sur les techniques mises en œuvre. Et quand elles le sont, il ne s'agit pas de toute la panoplie de l'éditeur mais de quelques outils bien précis, parfois éloignés de la notion même d'IA ! Parfois, ces études de cas sont celles de produits de startups comme **Luminoso** et sa solution d'analyse des textes, présentée chez IBM Watson IOT.



Les études de cas les plus documentées sont pour l'instant celles qui sont publiées comme résultats de travaux de recherche !

Si vous avez de belles études de cas de projets de deep learning dans de grandes entreprises françaises, n'hésitez pas à les partager en commentaires !

### Un large éventail de possibles

Une entreprise qui se lance dans un benchmark de l'IA large bande se rendra rapidement compte que les outils de l'IA peuvent apporter de la valeur dans un tas de contextes inattendus et extrêmement variés. Elle devra faire des choix et gérer ses priorités. Au bout du compte, c'est l'intérêt business qui primera, et celui de ses clients. Ce, d'autant plus que l'IA perd rapidement de son côté un peu magique quand on met le nez dedans et que l'on comprend ses différentes techniques.

Finalement, vouloir benchmarker l'IA aujourd'hui reviendrait à benchmarker "l'informatique" en général ! Vaste programme...

Cet article a été publié le 12 mai 2017 et édité en PDF le 17 mars 2024.  
(cc) Olivier Ezratty – "Opinions Libres" – <https://www.oezratty.net>