

Cloud Public et IA

QUELS LIENS POSSIBLES ?

LE CLOUD PUBLIC : UNE OBLIGATION POUR L'IA ?

BURGUN Vincent | Expertise Cloud Computing | 2018

Résumé

La thèse professionnelle suivante traite de l'impact et des possibilités du cloud public dans le développement et l'exploitation de l'IA en milieu professionnel. Elle s'appuie sur un travail de recherche et de différentes interviews réalisées auprès d'utilisateurs et de fournisseurs de solution cloud. La première partie établit une définition de ce qu'on appelle aujourd'hui l'intelligence artificielle, notamment avec la mise en avant du "Machine Learning" et du "Deep Learning". On s'aperçoit rapidement de l'importance de la donnée métiers et des besoins de ressources informatiques pour l'élaboration d'une IA efficace. La seconde partie est dédiée au cloud public. Ce dernier étant d'abord une offre de service, un certain nombre de critères doivent être considérés dont la consommation des ressources à la demande, une facturation à l'usage et d'autres caractéristiques. Trois types d'offres dédiées à l'IA sont exposées et illustrées par des exemples tirés du catalogue de produits des trois principaux fournisseurs de cloud public actuels. La dernière partie est une projection des offres précédemment traités dans une entreprise issue de la finance, un secteur particulièrement intéressé par l'IA. La conclusion met en avant les avantages potentiels du cloud pour le traitement de l'IA sous certaines conditions, incluant des points d'attention, comme la qualité et la confidentialité de la donnée ainsi que la bonne compréhension des technologies et de leurs capacités réelles à répondre à un besoin.

The following professional thesis discusses the impact and possibilities of the public cloud in the development and operation of AI in the workplace. It is based on research work and various interviews with users and cloud solution providers. The first part establishes a definition of what is now called artificial intelligence, in particular with the emphasis on "Machine Learning" and "Deep Learning". We quickly realize the importance of business data and IT resource requirements for the development of an effective AI. The second part is dedicated to the public cloud. Since the latter is first and foremost a service offer, a number of criteria must be considered, including on-demand resource consumption, pay-per-use billing and other characteristics. Three types of AI offerings are presented and illustrated with examples from the product catalogs of the three leading public cloud providers today. The last part is a projection of the offerings previously processed in a company from the financial sector, a sector particularly interested in AI. The conclusion highlights the potential benefits of the cloud for AI processing under certain conditions, including attention points, such as data quality and confidentiality, as well as a good understanding of the technologies and their real capabilities to meet a need.

Table des matières

1	Introduction	4
1.1	Démarche retenue	4
2	L'IA, une notion floue.....	6
2.1	Historique de l'IA.....	8
2.2	Les grandes technologies d'IA.....	11
2.2.1	Grands principes.....	12
2.3	État des lieux.....	15
2.3.1	Cas d'usage et besoin.....	15
2.4	Synthèse.....	17
3	Le cloud public, un environnement propice à l'IA ?.....	19
3.1	Le cloud public	19
3.1.1	Caractéristiques d'une offre de « cloud computing »	19
3.1.2	Caractéristiques et acteurs.....	22
3.2	Les différentes visions du cloud public	26
3.2.1	L'accès à des ressources techniques.....	26
3.2.2	Un modèle économique	26
3.2.3	Une approche collaborative.....	27
3.3	Les offres d'IA dans le cloud	27
3.3.1	Les offres pour clients experts sur l'IA.....	31
3.3.2	Les offres « clef en main ».....	39
3.3.3	Les offres assistées	42
4	IA et CLOUD, une projection possible chez un Asset Manager	49
4.1	Le métier d'Asset Manager.....	50
4.2	Un chatbot au service client	51
4.2.1	Le chatbot.....	51
4.2.2	Intérêts	51
4.2.3	Mise en application dans le cloud public.....	51
4.3	Un assistant à la prise de décision pour les gérants	53
4.3.1	Le gérant.....	53
4.3.2	Intérêts	53
4.3.3	Mise en application dans le cloud public.....	53

4.4	Le suivi du parcours client au marketing.....	56
4.4.1	Comprendre le raisonnement du client.....	56
4.4.2	Intérêts	56
4.4.3	Mise en application dans le cloud public.....	56
5	Conclusion	58
5.1	Des offres en devenir et un IAAS toujours présent	58
5.2	CLOUD et IA : une association qui fait mouche, mais avec laquelle il faut rester prudent	59
5.2.1	Un environnement optimisé.....	59
5.2.2	Une simplicité parfois dangereuse.....	60
5.2.3	Le côté obscur du cloud public.....	60
5.3	La formation et les compétences humaines : les compléments parfaits A l'IA dans le cloud public	61
6	Annexes.....	62
6.1	Interviews	62
6.1.1	AWS : Julien Simon : Technical Evengelist Artificial Intelligence.....	62
6.1.2	CDO d'un groupe bancaire.....	67
6.1.3	DataKlub : Tony Alibelli : CTO	71
6.1.4	Google : Jean-Luc Moisan : Responsable Avant-Vente Services Financier 75	
6.1.5	Nephelai : Romain Mangeret : Cofondateur.....	79
6.1.6	Orange : Julien Bermudez : Business Manager Cloud For Business.....	84
6.1.7	QuantMetry : Aurelia Nègre : Data Scientist.....	87
6.1.8	Quinten : Anna Nesvijevskaia Directeur Data Strategy et Thomas Schott Directeur technique	90
6.1.9	SESamm : Sylvain Forte : Cofondateur et CEO.....	94
6.1.10	Groupe TF1 : Philippe Toublant : Directeur Adjoint Data.....	97
6.1.11	Microsoft : Clement Le Roux : Cloud Solution Architect AI & Data.....	99
6.2	Autres sources.....	103
7	Glossaire.....	104

1 Introduction

La démocratisation du cloud et l'augmentation de son usage ont permis l'accès à de nouvelles technologies. La mise à disposition de ressources quasi illimitées et consommables au besoin facilite l'innovation et la création de nouveaux usages. Parmi ces nouvelles offres, l'IA (intelligence artificielle et assimilé) semble avoir profité énormément des avancées du cloud. Les entreprises sont de plus en plus demandeuses d'intégration d'IA dans leur métier et le cloud semble offrir un écosystème adapté et pertinent. Les principaux *providers** de cloud public ne cessent de communiquer sur leurs offres dédiées à l'IA (et assimilé) et aux capacités de leur infrastructure pour répondre aux problématiques associées. Dans ce contexte il serait intéressant d'identifier les liens entre clouds et IA afin de mesurer cette pertinence à travers une projection sur une entreprise d'Asset Management.

1.1 DÉMARCHE RETENUE

Afin d'être le plus pertinent possible dans mes propos, j'ai décidé de baser mon travail sur une série d'interviews. J'ai pu échanger avec des fournisseurs et des utilisateurs de cloud ainsi qu'avec des sociétés utilisant ou proposant des solutions d'IA. La plupart de mes interlocuteurs sont des professionnels évoluant dans le domaine financier. En effet, ce secteur est particulièrement sensible au sujet d'IA. De par leur activité, les acteurs concernés manipulent historiquement de la donnée, et comme nous le verrons plus tard dans ce document, la donnée reste la matière première de l'IA.

Vous pouvez retrouver en annexe de ce document, l'intégralité des interviews réalisés. Je tiens d'ailleurs dès à présent à remercier les participants pour l'aide précieuse qu'ils m'ont apportée dans la rédaction de cette thèse professionnelle et de l'intérêt des échanges que nous avons pu avoir.

Le sujet sera traité de la façon suivante :

- Une première partie chargée de faire un état des lieux de l'IA afin d'essayer d'en déterminer les principes et les besoins
- Une seconde partie chargée de définir le cloud public et d'identifier les offres en mesure de traiter les besoins de l'IA
- Une troisième partie chargée d'illustrer les applications possibles de l'IA dans une architecture cloud en prenant l'exemple d'une société d'Asset Management

Je terminerai sur une conclusion afin de résumer les points importants du sujet.

* (Les mots en italique suivi d'une étoile sont définis dans le glossaire présent en annexe du document)

J'ai volontairement choisi d'étudier uniquement le cas du cloud public que je présenterai dans la seconde partie de ce document. J'ai effectivement trouvé plus pertinent de concentrer mes recherches autour des principaux fournisseurs de ce marché. D'ailleurs pour la majorité de mes interlocuteurs, le cloud était associé principalement au cloud public. C'est également le type de cloud qui connaît un très haut niveau de croissance et un intérêt grandissant de la part des entreprises.

La demande des entreprises pour du cloud public s'oriente vers des services intégrés IaaS et PaaS. Une aubaine pour AWS, Microsoft, Alibaba et consorts.

Fortement dynamique, le marché mondial des services de cloud public pourrait peser 206,2 milliards de dollars en 2019, contre 175,8 milliards en 2018, rapporte Gartner.

Les prévisions de la société d'études sont moins élevées que celles publiées au printemps dernier. Malgré tout, le secteur affiche toujours une solide croissance à deux chiffres (**+21% attendus en 2018**, +17,3% en 2019).

1 : source : <https://www.silicon.fr/cloud-public-croissance-2019-219733.html>

Ce document ne se prétend pas être une thèse spécialisée sur l'IA traitant en profondeur des réalités mathématiques et scientifiques qui peuvent en découler. Les explications associées sont donc volontairement simplifiées afin de rester cohérent avec l'objectif principal du document : faire un état des lieux de l'usage du cloud computing pour traiter des problématiques d'IA et savoir si le cloud est aujourd'hui un facteur clef pour la mise en application d'IA dans les entreprises.

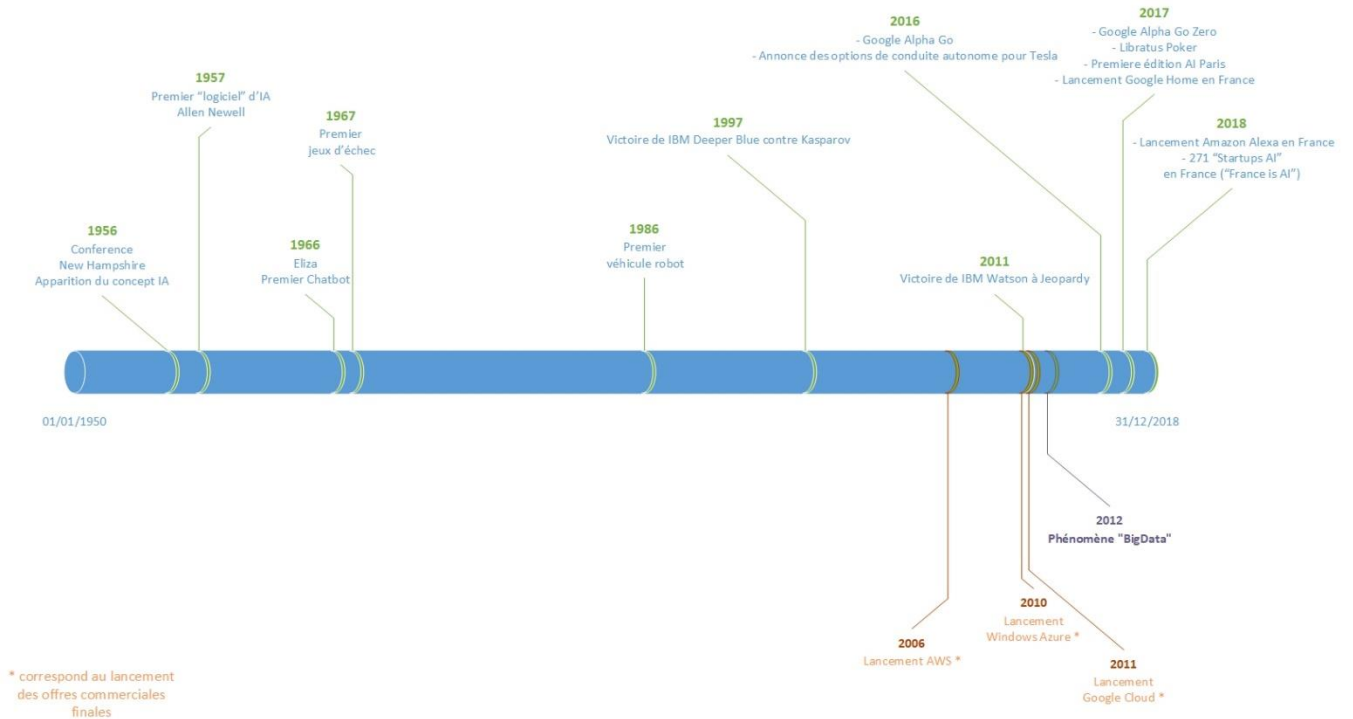
- L'IA ce sont de **nouveaux services** offrant aux entreprises de nouvelles méthodologies de travail
- L'IA est le domaine scientifique qui traite de l'étude, la conception et la mise en œuvre de machines « intelligentes » (machines prenant le sens de système automatique capable de traiter l'information) (*Jean-Claude Heudin : Comprendre le Deep Learning* »)

Il y'a une volonté très forte d'imiter le cerveau humain. Ce dernier étant toujours à l'étude de nos jours, il reste possible que l'IA évolue en parallèle des découvertes que l'on pourrait faire sur le cerveau.

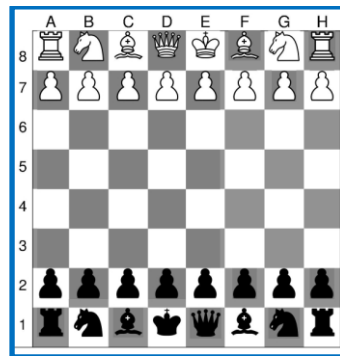
Partant de ces observations, j'ai souhaité, à travers les chapitres suivants, rendre la notion d'IA moins floue pour en proposer une définition moderne. L'objectif est de la confronter par la suite aux offres du cloud public.

2.1 HISTORIQUE DE L'IA

Avant d'aller plus loin sur ce qu'est l'IA et quels en sont les technologies et usages actuels, il me paraissait intéressant d'en faire un rapide historique. Contrairement à ce que l'on pourrait croire, l'IA n'est pas « une nouveauté ». Je vous propose de retracer chronologiquement les principaux épisodes associés. J'y ai inclus des événements liés au cloud computing pour essayer d'identifier un axe d'évolution commun.

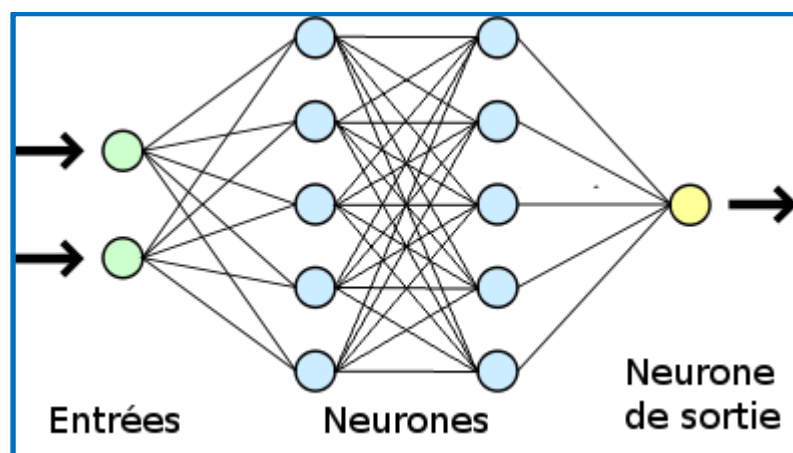


Le concept d'IA d'un point vu scientifique est né pour la première fois lors d'une conférence entre chercheurs en 1956. L'idée était d'essayer de représenter mathématiquement « un esprit ». L'approche a d'abord été nommée « Symbolique », en partant du principe que le cerveau humain est une machine à traiter les symboles. Des années 60 à 70, beaucoup d'algorithmes vont découler de cette façon de traiter l'IA, notamment les premiers algorithmes pouvant permettre à une machine de jouer aux échecs.



3 : source : *Thomas Cambrai - L'intelligence artificielle expliquée*

Un autre courant va émerger, celui du numérique, censé être plus proche du fonctionnement réel du cerveau humain. C'est ce courant qui va donner lieu aux « réseaux de neurones »



4 : source : <https://charlesen.fr/reseaux-de-neurones-en-javascript-avec-brain-js/>

Malgré ces avancées, la fin des années 70 connaît le début de « l'hiver de l'IA ». En effet, les recherches coûtent cher et les résultats ne sont pas au rendez-vous.

Ce qu'il est important de comprendre, c'est que malgré cet arrêt des recherches, **beaucoup des briques théoriques utilisées dans l'IA d'aujourd'hui ont été découvertes à cette époque.**

Il faudra attendre néanmoins le milieu des années 90 pour à nouveau voir l'IA sur le devant de la scène, avec notamment les démonstrations d'IBM (acteur du cloud aujourd'hui) et la fameuse victoire de la machine contre le champion mondial des échecs, Kasparov (1997).

Mais encore, une fois, le marketing va prendre le pas sur les faits, et les attentes sont encore une fois trop fortes au vu des possibilités offertes (et surtout des coûts nécessaires)

Beaucoup de mes intervenants m'ont indiqué avoir vu un vrai regain de l'IA suite à l'avènement du « BigData* ». Si les principes mathématiques et algorithmiques pour mettre en œuvre une IA étaient disponibles depuis longtemps, les besoins en données sources et en puissance de traitement n'étaient pas couverts. Le « BigData » a permis la mise en avant de technologies de gestion de la donnée innovante et a levé une partie des freins techniques nécessaires à l'IA pour absorber de la donnée.

Ce qui est intéressant, c'est que le « BigData » a été l'un des cas d'usage pour certaines entreprises de bascule dans le cloud public. En effet, la majorité des clouds providers publics se sont positionnés sur ce marché, notamment grâce à leurs offres de stockage objets et leur infrastructure « IaaS » (Infrastructure As A Service) à même de traiter de gros volumes d'informations.

L'adoption des outils Cloud BI a augmenté de 50% en trois ans

Entre 2013 et 2016, l'adoption des outils Business Intelligence sur le cloud a augmenté de 50%. Selon BARC Reserach, alors que 29% d'entreprises les utilisaient en 2013, le **pourcentage est passé à 43% en 2016.**

5 : source : <https://www.lebigdata.fr/chiffres-big-data>



Cloud et Big Data :
une combinaison gagnante

Une récente enquête conduite par GigaSpaces a révélé que 80 % des responsables informatiques pour qui le traitement Big Data est important envisagent de transférer leurs systèmes analytiques sur un ou plusieurs modèles de distribution Cloud.⁴

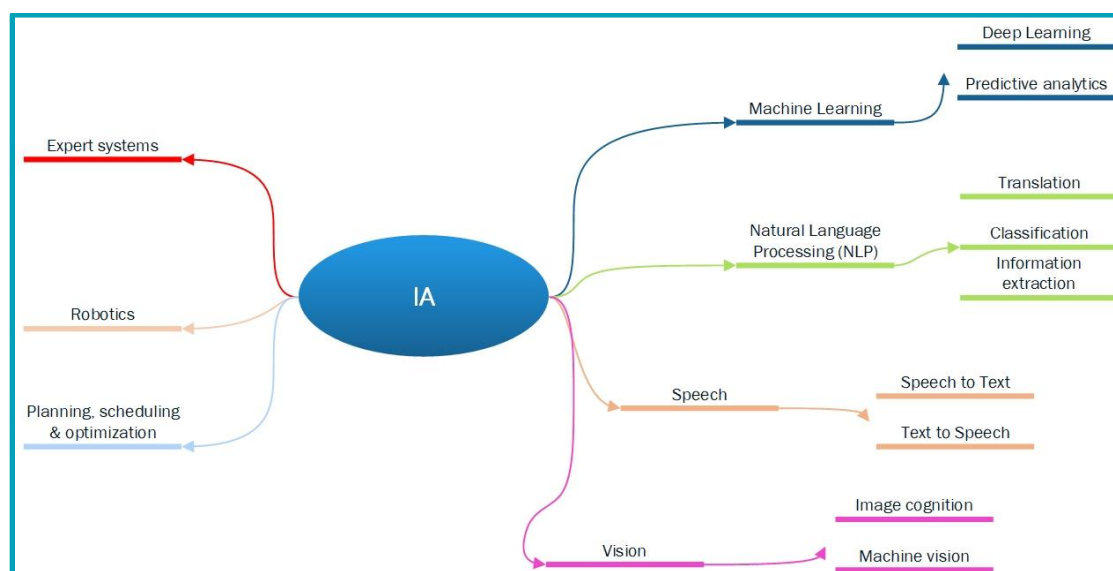
6 : source : <https://www.intel.fr/content/www/fr/fr/big-data/big-data-cloud-technologies-brief.html> (2013)

À partir de 2012, on note la mise en avant de plus en plus forte des cas d'usage de l'IA, dont beaucoup sont portés par des acteurs du cloud computing comme Google, AWS, et dans une moindre mesure Microsoft notamment avec leur mise sur le marché de leur assistant connecté.

2.2 LES GRANDES TECHNOLOGIES D'IA

Après ce rapide historique de l'IA et de son évolution, il est intéressant de se focaliser sur les technologies utilisées aujourd'hui. Nous pourrions ainsi les comparer aux propositions des clouds providers publics pour savoir s'ils peuvent facilement ou non s'y appliquer.

L'IA se concrétise par différentes possibilités. Voici un schéma qui résume les principaux domaines d'activité de l'IA.



7 : source : Thomas Cambrai - L'intelligence artificielle expliquée

Pour la suite du document, nous parlerons essentiellement de la partie droite du schéma regroupant :

- L'analyse et le traitement du langage (traduction, classification, information et interaction)
- L'analyse et le traitement de l'image (reconnaissance, réalité augmentée)

Ce sont en effet ces besoins qui m'ont été le plus souvent rapportés lors de mes interviews et qui ont révélé un intérêt certain pour :

- Le traitement du langage humain (allant de sa retransmission orale ou écrite (*Text to Speech*/Speech to Text**) à l'analyse sémantique des mots (*Natural Language Processing**))
- Le Machine Learning
- Le Deep Learning.

J'ai été étonné que tout le monde ne s'accorde pas sur la même définition de certains de ces termes, notamment pour le « Machine learning » et le « Deep learning »

Je vous propose donc maintenant de faire un focus plus particulier sur ces derniers.

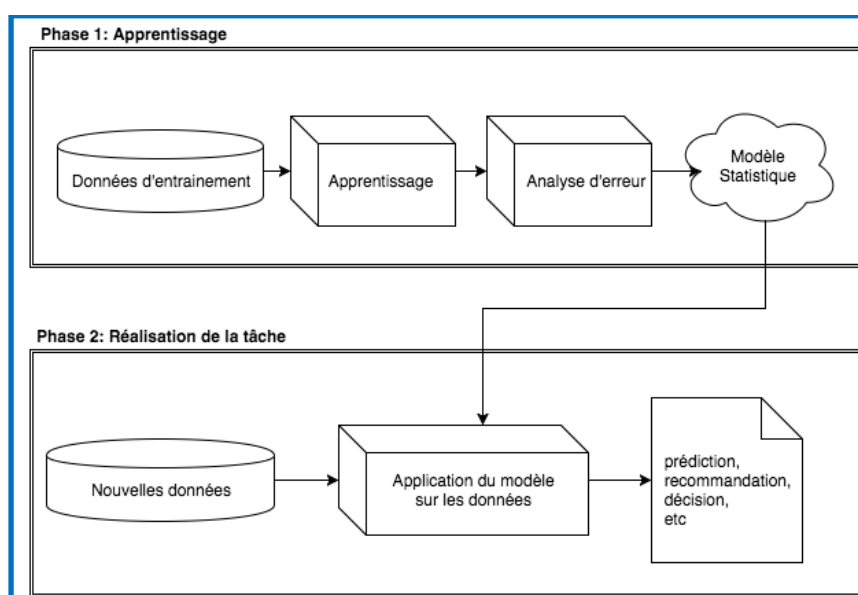
2.2.1 Grands principes

Qu'il s'agisse du « Machine learning » ou du « Deep learning », le but est de proposer des systèmes en mesure d'apprendre un raisonnement en fonction d'une quantité de données en entrée. On distingue toujours deux phases

- La phase d'apprentissage. On découpe un jeu de données connues en deux parties.
 - La première partie est utilisée pour entraîner la machine et obtenir un modèle de traitement. On peut alors indiquer à la machine les caractères discriminants (apprentissage supervisé) ou laisser la machine détecter elle-même les éléments différenciant dans l'information (apprentissage non supervisé)
 - La seconde partie est utilisée pour évaluer le modèle obtenu et indiquer à la machine les corrections éventuelles à réaliser pour proposer un modèle opérationnel.

À noter qu'on procède souvent à de nombreuses phases d'apprentissages, l'obtention d'un modèle efficace réclame souvent plusieurs itérations.

- La phase d'exploitation : le système une fois entraîné, traite de nouvelles données d'entrée non connues. Le modèle obtenu en phase d'apprentissage fournit désormais une information de sortie exploitable par l'entreprise.



8 : source : <https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning/4011858-quest-ce-que-le-machine-learning>

2.2.1.1 Machine Learning

Après avoir croisé les réponses qui m'ont été faites et la lecture d'ouvrages spécialisés (références disponibles en annexe), voici quelques détails sur ce qu'on appelle le « Machine Learning »

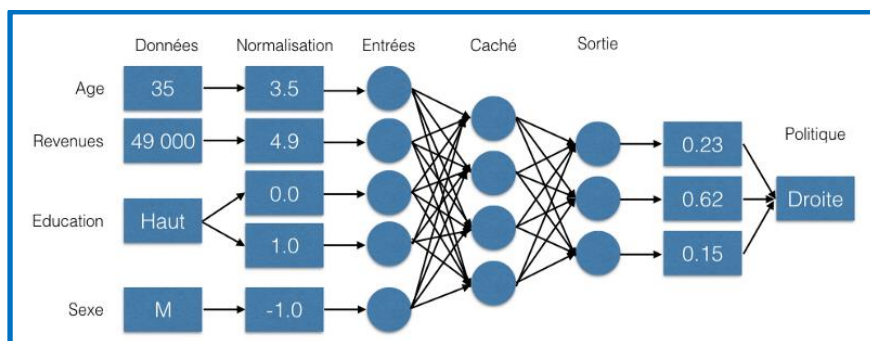
Il s'agit de la mise en œuvre de systèmes capables d'apprendre des comportements sans être explicitement programmés grâce à des données travaillées. On utilise pour cela des méthodes très variées allant de simple régression linéaire à des arbres de décision complexe (ex *Random Forest**). La réponse apportée est toujours de type probabilistique.

Par abus de langage et confusion, on parle souvent de « Machine Learning » ou « ML », uniquement dans le cas d'usage de traitement de données structurées. Dans ce type d'usage, il n'est pas toujours nécessaire d'avoir énormément d'informations en entrée des algorithmes utilisés. Une fois les éléments de corrélations trouvés, les performances ont tendance à stagner, peu importe le nombre de sources analysées. La sortie est le plus souvent un résultat numérique. Si ces exemples sont effectivement des cas de « Machine Learning », ce ne sont pas les seules applications du ML comme nous allons le voir dans la partie suivante.

2.2.1.2 Deep Learning.

Le « Deep Learning » ou apprentissage profond est sans doute le terme le plus mis en avant récemment sur l'IA. Pourtant, c'est celui sur lequel j'ai rencontré le plus de divergences dans sa classification durant les interviews. Bien que cela ne fasse pas l'unanimité, **le « Deep Learning » est un sous-ensemble du « Machine Learning »**. Il repose exactement sur les mêmes principes, mais s'appuie sur ce qu'on appelle un réseau de neurones pour offrir des capacités de traitement plus proche de celui du cerveau humain et plus efficace dans certains usages, notamment dans la manipulation de données non structurées. Le « ML » englobant l'apprentissage profond, on ne peut donc pas résumer le « ML » au cas d'usage vu précédemment.

Un réseau de neurones est un système dans lequel plusieurs caractéristiques d'un jeu de données sont présentées à l'entrée de microprogrammes interconnectés entre eux sous forme de couches.



9: source : https://fr.wikipedia.org/wiki/R%C3%A9seau_de_neurones_artificiels

Comme le « Deep Learning » est une forme de « Machine Learning », il nécessite une phase d'apprentissage, mais tire partie d'une plus grande quantité de données en entrée. Plus le réseau de neurones dispose de données d'entraînement et d'historique, plus il est à même de fournir un résultat avec une marge d'erreur minime. Sa mise en œuvre peut donc nécessiter des besoins de stockage et des capacités de traitement important.

À noter que le « Deep Learning » peut être décliné sous une autre variante, appelée « Reinforcement learning », même si à l'origine ce type de système pouvait s'appliquer sans réseau de neurones, il tire désormais également partie de ces architectures. Dans le « Reinforcement learning », l'apprentissage se fait par l'observation et l'essai avec un système de récompenses/punitions. On peut par exemple l'exploiter pour créer une IA de conduite automatique.



Figure 10 : photo prise lors de l'évènement "AI Essentials" organisé par Microsoft le 26/09/2018

Ce qui est intéressant à noter c'est que les réseaux neuronaux ont été conceptuellement imaginés avant certains algorithmes de machine Learning comme la classification par « Forêt d'arbres décisionnels » mis en avant en 2001. Bien que le « Deep Learning » soit un sous-ensemble du « Machine Learning », il repose sur des concepts plus anciens. Les réseaux neuronaux auront donc mis du temps avant d'être mis en œuvre efficacement ; cela encore une fois faute de moyens techniques suffisamment performants à l'époque et de volumes de données suffisants pour qu'ils se montrent efficaces.

Il est également nécessaire de parler des problèmes d'« explicabilité ». Un réseau neuronal évoluant de lui-même, s'il est possible d'expliquer son fonctionnement, il est actuellement impossible d'expliquer son raisonnement. En effet, son comportement va évoluer au fur à mesure de la quantité d'informations qu'il peut avoir à disposition pour s'entraîner. Dès lors il n'est plus possible pour un humain de pouvoir justifier et comprendre pourquoi un réseau neuronal a donné telles informations en sortie. Cela peut poser problème dans certains domaines où la loi impose des conditions sur le traitement de l'information (ex : refus de crédit) ou que la donnée de sortie est tellement critique qu'on ne peut laisser la prise de décision à une IA dont on n'est pas capable de garantir la cohérence de sa réponse (ex : déclenchement d'une arme)

2.3 ÉTAT DES LIEUX

2.3.1 Cas d'usage et besoin

Sur les cas d'usages rencontrés à travers les interviews réalisées, certains sujets reviennent régulièrement pour justifier de la mise en place d'un système d'IA.

2.3.1.1 Traitement du langage humain et de la voix

L'un des premiers est le « NLP » (*Natural Language Processing**) pour le traitement du langage humain. Cela va des besoins de traduction, à l'interprétation des sentiments en passant par la reconnaissance vocale et la transformation d'élément écrit en élément oral (*Text To Speech**).

Certains de mes interlocuteurs s'accordent pour dire que le futur de l'informatique sera l'usage de la voix pour interagir avec nos équipements électroniques et nos applications. La voix est en effet le vecteur de communication principal de l'être humain, il paraît donc logique que la manipulation des outils informatiques se fasse par ce biais. L'« IA » est très fortement attendue sur ce sujet et beaucoup d'avancées ont été faites dans le domaine du NLP.

50% des requêtes BI seront effectuées vocalement d'ici 2020

Selon Gartner, d'ici 2020, 50% des requêtes de Business Intelligence seront effectuées au son de la voix d'ici 2020 grâce à la **reconnaissance vocale et au traitement naturel du langage**. Ces technologies liées à l'intelligence artificielle vont permettre à tout un chacun d'utiliser les outils de Business Intelligence pour consulter les données, sans avoir de compétences techniques.

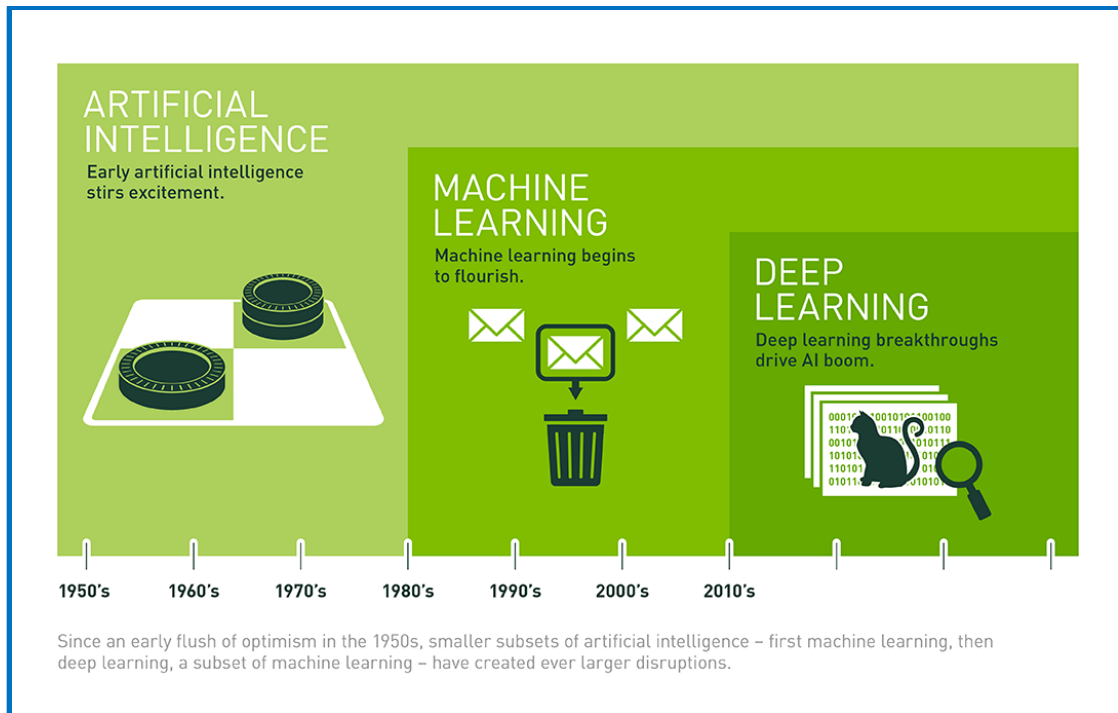
11 : source : <https://www.lebigdata.fr/chiffres-big-data>

2.3.1.2 Traitement de documents

L'autre besoin fort exprimé à travers l'interview est autour de la classification de documents et reconnaissance d'image. La multiplication des données et l'intégration de plus en plus d'éléments non structurés dans les informations traitées par les entreprises poussent à développer des systèmes d'IA spécialisés.

À titre d'exemple, un service de stockage de photo comme « Google Photos » exploite aujourd'hui 64 filtres d'analyse pour pouvoir traiter une image et la classer.

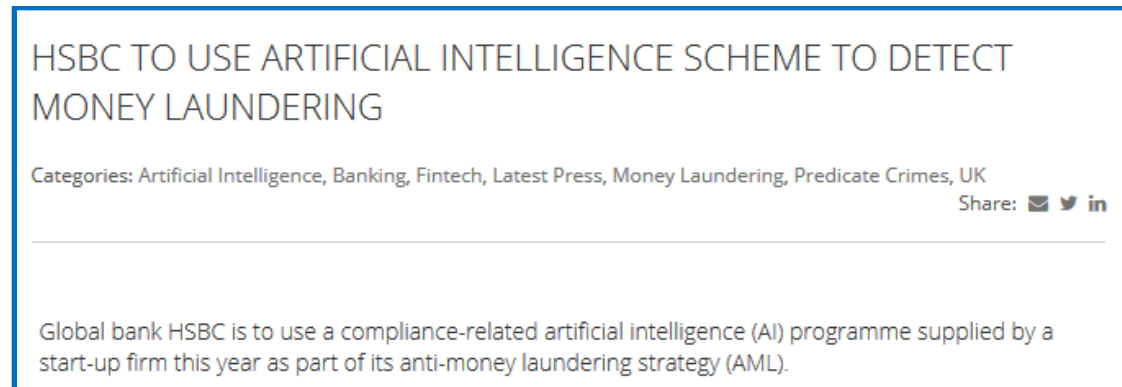
Le besoin d'automatisation de certaines tâches est également souvent exprimé notamment dans le milieu bancaire. On parle alors de RPA « pour Robot Process Automation ». Ce sujet est particulier, car si certains l'associent à l'IA, le fait d'automatiser une tâche n'implique pas forcément un apprentissage ou le besoin de raisonner. **Une application codée pour réaliser une action répétitive en fonction de critères explicitement définis ne semble plus faire partie de l'IA « moderne » telle qu'on essaye de la mettre en œuvre aujourd'hui.**



12 : source : <https://www.ledigitalab.com/fabrique/intelligence-artificielle-machine-learning-deep-learning-kezako/>

2.3.1.3 L'aide à la décision

Le milieu financier investit également dans l'IA pour l'identification de la fraude. L'IA permet l'analyse comportementale et la mise en avant de corrélation pouvant assister les services spécialisés des banques.



13 : source : <https://kyc360.com/news/hsbc-use-artificial-intelligence-ai-scheme-detect-money-laundering/> (AML = lutte contre le blanchiment d'argent)

L'aide à la prise de décision est d'ailleurs un usage souvent exprimé pour la justification de la création d'une IA. C'est notamment un secteur sur lequel se sont positionnés beaucoup de startups, et plus particulièrement les fintech (startups qui utilisent les nouvelles technologies à destination de la finance).

2.3.1.4 Le parcours client

Les services marketing sont également très intéressés par la mise en place d'une IA pour connaître au mieux le parcours de leur client et identifier les meilleures stratégies commerciales en fonction des cibles à atteindre.

2.4 SYNTHÈSE

Qu'il s'agisse du traitement du langage, de la reconnaissance d'images ou de l'aide à la décision, l'IA est **aujourd'hui principalement déployée à travers des systèmes de « Machine Learning »**. En fonction des cas d'usage et des domaines, les algorithmes déployés sont plus ou moins complexes et nécessitent plus ou moins de données avec des besoins de calculs plus ou moins élevés. Le « Deep Learning » représente une branche avancée du « Machine Learning » qui n'a pu réellement se montrer pertinent qu'à partir du moment où la technologie a permis d'offrir suffisamment de puissance pour la mise en œuvre de réseaux de neurones.

En faisant le lien avec l'histoire de l'IA et ses technologies actuelles, on pourrait adopter la définition suivante

L'« IA » (actuelle) est l'interprétation d'un volume important de données dans le but de proposer une réponse approximative à une question. Pour cela, on applique des principes mathématiques traduits au sein d'algorithmes chargés de se rapprocher au mieux du fonctionnement du cerveau humain. Le défi est d'avoir la marge d'erreur la plus faible possible.

Cette proposition est un parti pris de ma part. Elle peut exclure une partie des technologies « d'IA » vues précédemment comme l'automatisation. Néanmoins elle couvre les usages modernes de l'IA et la très forte appétence du marché pour « le Machine Learning »

C'est principalement sur cette définition accompagnée des points vus précédemment que je vais analyser la possibilité du cloud public pour la mise en œuvre d'IA.

On distingue déjà deux composants importants qu'il faudra mettre en parallèle du cloud computing.

- **La donnée** : c'est aujourd'hui la matière première de l'IA
- **Le traitement** : l'IA nécessite un environnement d'exécution pour fonctionner.

3 Le cloud public, un environnement propice à l'IA ?

Maintenant que l'on a établi un état des lieux de l'IA et une définition associée, je vais me focaliser sur le cloud public. Celui-ci a-t-il un impact sur le déploiement de l'IA, permet-il la mise en œuvre des technologies de « Machine Learning » vu précédemment, si oui, sous quelles formes et avec quelles conditions.

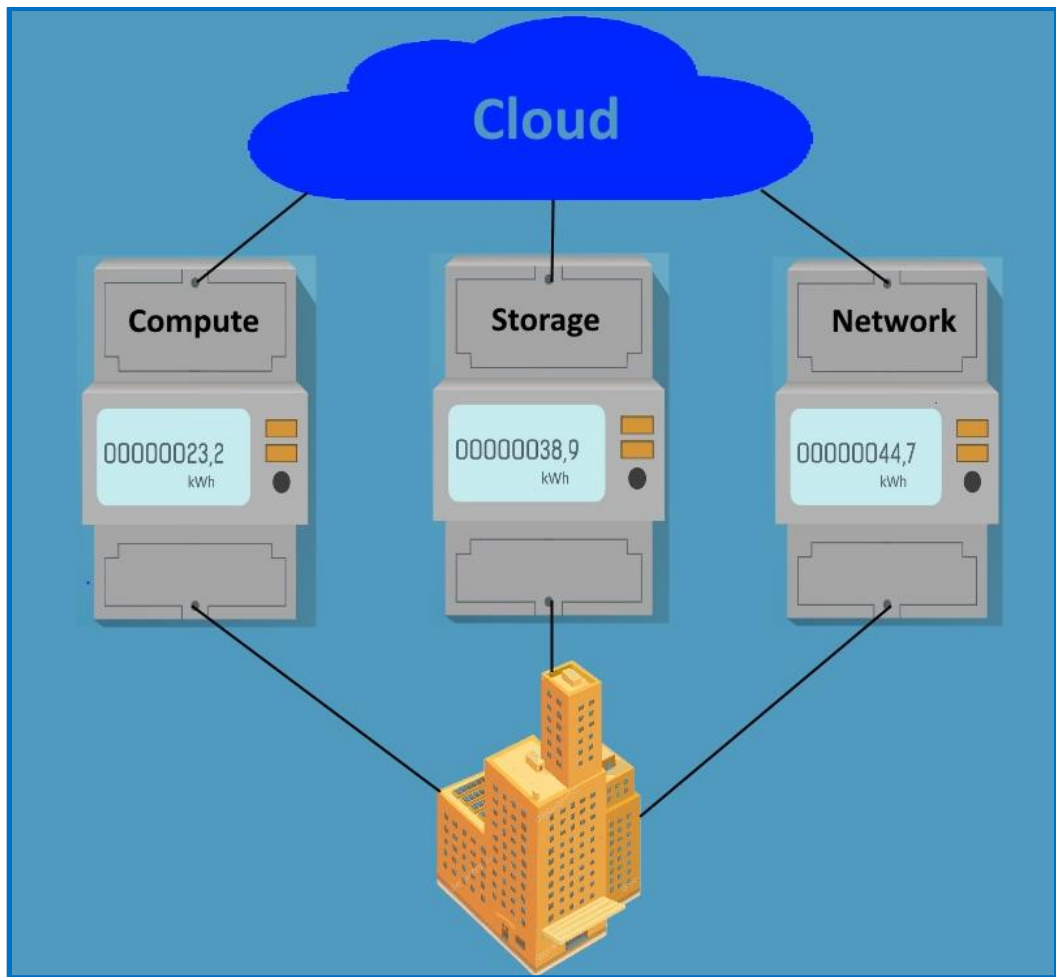
3.1 LE CLOUD PUBLIC

Il me semblait intéressant de définir ce qu'est le cloud public avant d'analyser les offres d'IA associées.

3.1.1 Caractéristiques d'une offre de « cloud computing »

Le cloud, peu importe sa forme (privée, public, hybride ou communautaire) est avant tout **une offre de services** de la part d'un fournisseur (interne à l'entreprise dans le cas du cloud privé). C'est une méthode de consommation de l'outil informatique qui respecte les caractéristiques suivantes.

- **Une offre en libre-service.** Le « cloud computing » est adressable par l'utilisateur lui-même qui est totalement autonome sur le déploiement de la solution choisie et son usage. Derrière ce point se cachent l'automatisation et l'industrialisation de l'architecture technique.
- **Une offre accessible sur l'ensemble d'un réseau** (internet pour le cloud public). Le « cloud computing » est adressable depuis plusieurs points via plusieurs périphériques. Derrière ce fait se cachent les interconnexions de l'ensemble des datacentres d'un fournisseur et la capacité du cloud à fournir des services à très hautes disponibilités.
- **Une offre basée sur la mutualisation des ressources.** Le « cloud computing » repose sur des équipements partagés entre plusieurs clients ou la segmentation est assurée par des composants logiciels. Derrière ce point se cache le modèle économique du cloud pour les fournisseurs : plus ils peuvent mutualiser leur architecture, plus ils gagnent de l'argent.
- **Une offre qui propose de l'élasticité sur ses services.** Le « cloud computing » ne doit pas afficher de limites visibles pour ses clients. Si l'on a besoin de plus de ressources ou moins de ressources, on peut ajuster sa consommation. Cette élasticité doit être la plus immédiate possible. Derrière ce point se cache le modèle économique de l'utilisateur du cloud. Vous payez ce que vous consommez réellement.
- **Une offre qui propose une mesure précise de la consommation de ses services.** Le « cloud computing » doit pouvoir afficher de façon claire et précise la consommation des ressources qui sont exploitées par le client. Derrière ce point se cache l'un des challenges les plus difficiles : comprendre les offres et les modèles de facturation pour suivre le bon usage de son cloud et choisir les meilleures options, au bon moment, compatibles avec son besoin et ses usages.



14 : Illustration imagée du cloud (site personnel : www.buzzword.tech)

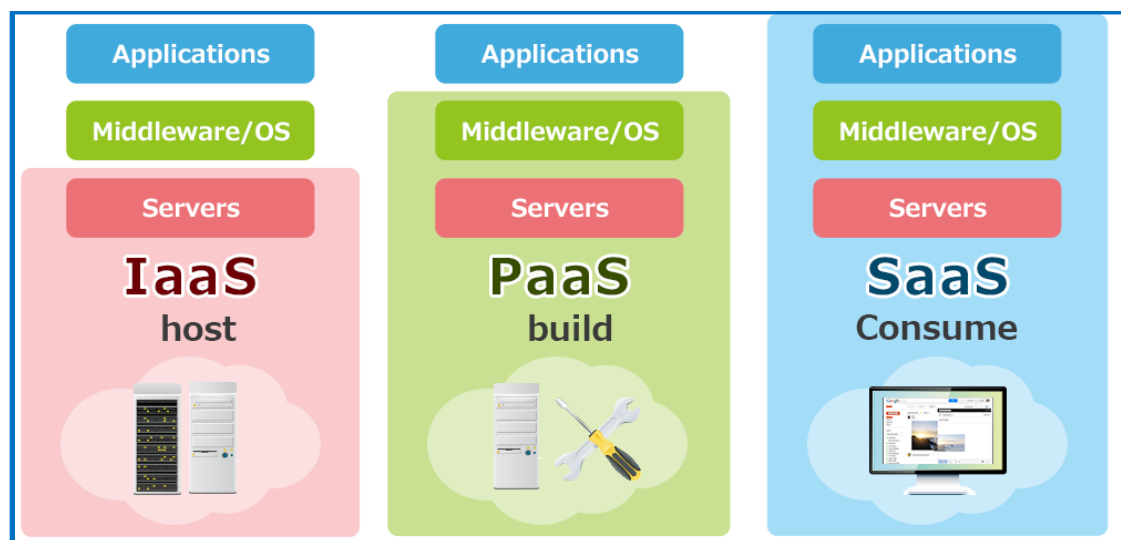
Une offre de cloud computing peut se proposer principalement sous trois formes

- IaaS : « Infrastructure As A Service ». C'est la forme la plus répandue et la moins assistée. L'utilisateur provisionne des ressources serveur, principalement sous forme de machines virtuelles (VMs*) qu'il peut associer à une configuration réseau spécifique, à des ressources de stockages ou à d'autres services additionnels (système de répartition de charges, de protections d'accès...) **Dans le cas du « IaaS » l'utilisateur n'a pas à se préoccuper de la couche physique** qui héberge ses serveurs. Il peut exploiter des images de systèmes d'exploitation déjà prêtes à l'emploi proposées par son fournisseur pour accélérer ses déploiements. L'utilisateur reste néanmoins responsable de l'usage de ses serveurs et des outils/applications qu'ils déploient par-dessus. La plupart du temps, le modèle de facturation retenu est à la seconde d'utilisations des machines virtuelles provisionnées et proportionnelles à la puissance des configurations demandées (nombre de processeurs, quantité de mémoire...)

- PaaS : « Plateforme As A Service ». C'est la forme qui se développe le plus actuellement. Elle propose un niveau d'abstraction complémentaire au « IaaS ». L'utilisateur exploite une solution de traitement et de développement prêt à l'emploi. Il n'a plus à provisionner de ressources, mais à développer directement son application sur la plateforme « PaaS ». **C'est le fournisseur qui prend en charge le provisionnement et l'administration totale des ressources sous-jacentes.** Les modèles de facturations associés sont plus complexes que pour le IaaS et peuvent prendre en compte plusieurs facteurs, comme le nombre de fois où le code de l'utilisateur est exécuté dans la plateforme « PaaS » et/ou les temps de traitements.
- SaaS : « Software As A Service ». C'est la forme qui connaît le plus gros catalogue d'offres aujourd'hui. Il s'agit pour l'utilisateur d'accéder et d'utiliser une solution logicielle prêt à l'emploi. **Le fournisseur s'occupe de l'ensemble de la chaîne, depuis l'hébergement de la solution avec le maintien en condition opérationnel, jusqu'au développement de l'application** et sa mise à disposition de l'utilisateur.

Ces trois formes sont liées entre elles, car elles doivent toutes respecter les caractéristiques d'une offre cloud. On observe ainsi le plus souvent des offres « SaaS » qui reposent sur des offres « PaaS », elles-mêmes exécutées sur une offre « IaaS ». Cette connexion est rarement visible de l'utilisateur. Ce dernier choisit à quels niveaux il souhaite intervenir et quel niveau de responsabilité et d'assistance il délègue à son fournisseur.

D'autres formes d'offres de cloud public sont disponibles comme le « CaaS » pour « Conteneur As A Service », mais il s'agit le plus souvent d'une sous famille ou d'un mélange des formes vues précédemment.



15: source : <https://medium.com/@Albihany/true-cloud-story-about-iaas-paas-saas-47cfea883271>

3.1.2 Caractéristiques et acteurs

Le « cloud public » est caractérisé par le fait que son fournisseur propose ses services à toutes entreprises et/ou particuliers qui le souhaitent, sous conditions de l'acceptation d'un contrat d'utilisation. Les offres sont toutes accessibles à **travers le réseau internet**.

Un fournisseur ou « provider » de « cloud public » propose donc des offres de « cloud computing » via la ou les formes souhaitées (IaaS, PaaS, SaaS) avec un modèle de facturation de son choix.

Le marché est aujourd'hui principalement occupé par trois entreprises américaines

- Amazon Web Service ou « AWS ». Leader actuel sur plusieurs segments. Fournisseur historique de solution IaaS, AWS est un acteur majeur du cloud public.



- Microsoft Azure. Arrivé un peu plus tard sur le cloud public, il dispose néanmoins de parts de marché importantes, notamment à travers son offre bureautique en SaaS (Office 365)



- Google Cloud Platform ou « GCP ». Le géant de la recherche sur Internet privilégie les offres « serverless », c'est-à-dire avec une abstraction totale de la couche physique abritant ses solutions.



Comme on peut l'observer sur le « Magic Quadrant » du « Gartner », d'autres fournisseurs sont disponibles sur le marché du « Cloud Public ». (À noter que le classement proposé peut varier si l'on cible une forme d'offre particulière (SaaS, PaaS, IaaS...))



16 : Source : Gartner - Mai 2018

En France, on trouve également des entreprises proposant des offres de cloud public comme Orange.

Quelques offres à la une :

Flexible Engine

L'offre de cloud public OpenStack d'Orange. Configurez votre infrastructure, sécurisez vos données, gérez votre sécurité, boostez les performances de votre réseau et pilotez votre cloud avec les Open APIs de Flexible Engine.

[Découvrir l'offre](#)

Flexible Computing Advanced

Combinez les avantages du cloud public et du cloud privé grâce au cloud privé virtuel. Créez et pilotez votre infrastructure cloud en toute sécurité, via les API VMware vCloud Director. Une solution modulaire et facturable à l'usage.

[Découvrir l'offre](#)

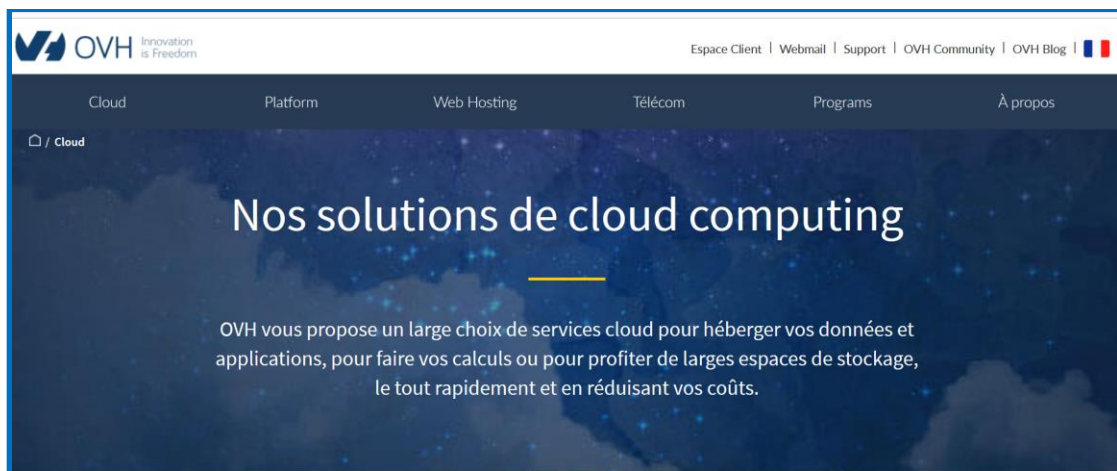
Flexible Storage

Sauvegardez, synchronisez et partagez vos données grâce au cloud. Adaptez votre solution de sauvegarde à l'usage de vos collaborateurs et maîtrisez votre budget grâce à une facturation à l'usage ou au forfait.

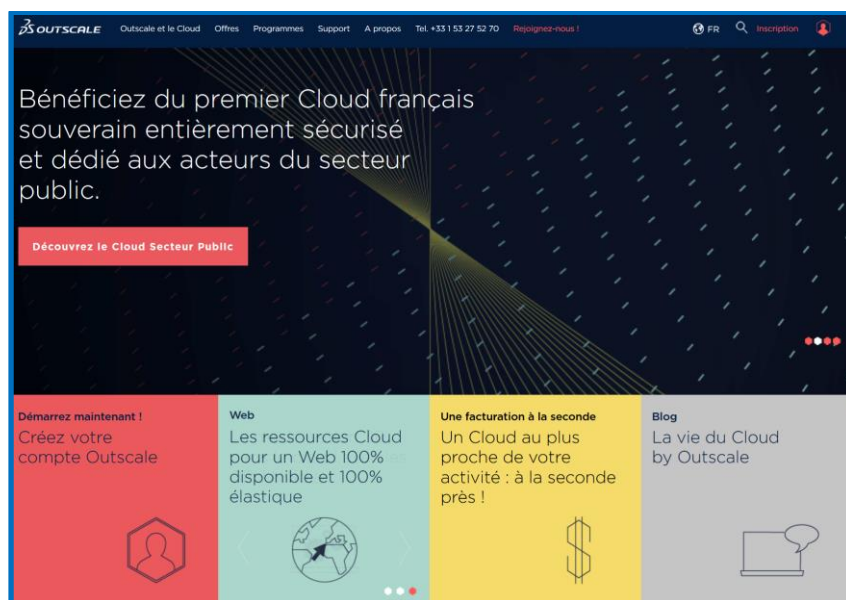
[Découvrir l'offre](#)

17 : source : <https://cloud.orange-business.com/fr/>

On retrouve également « OVH » ou « Outscale »



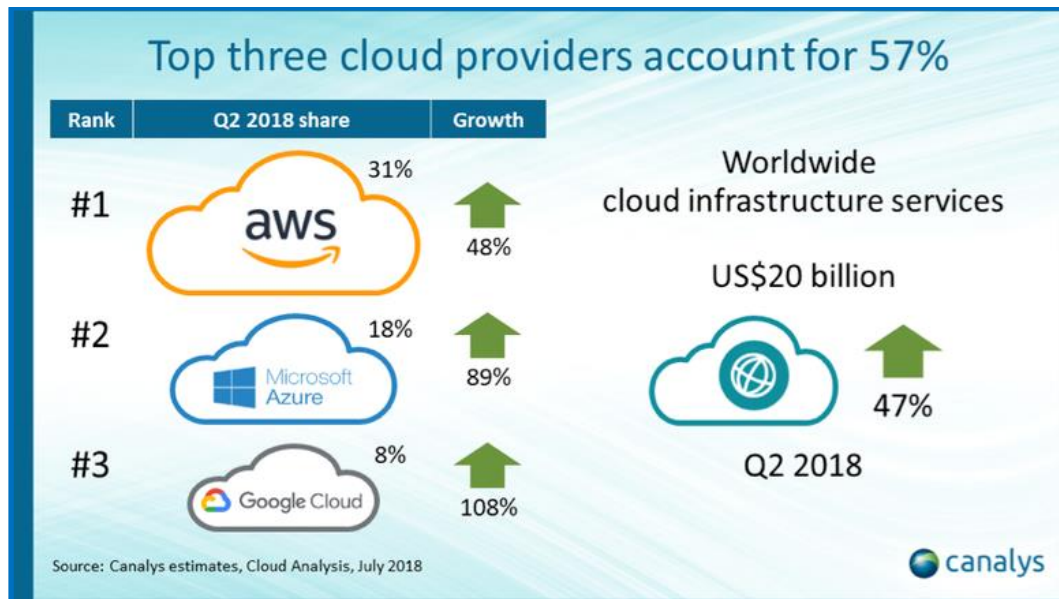
18 : source : <https://www.ovh.com/fr/cloud/>



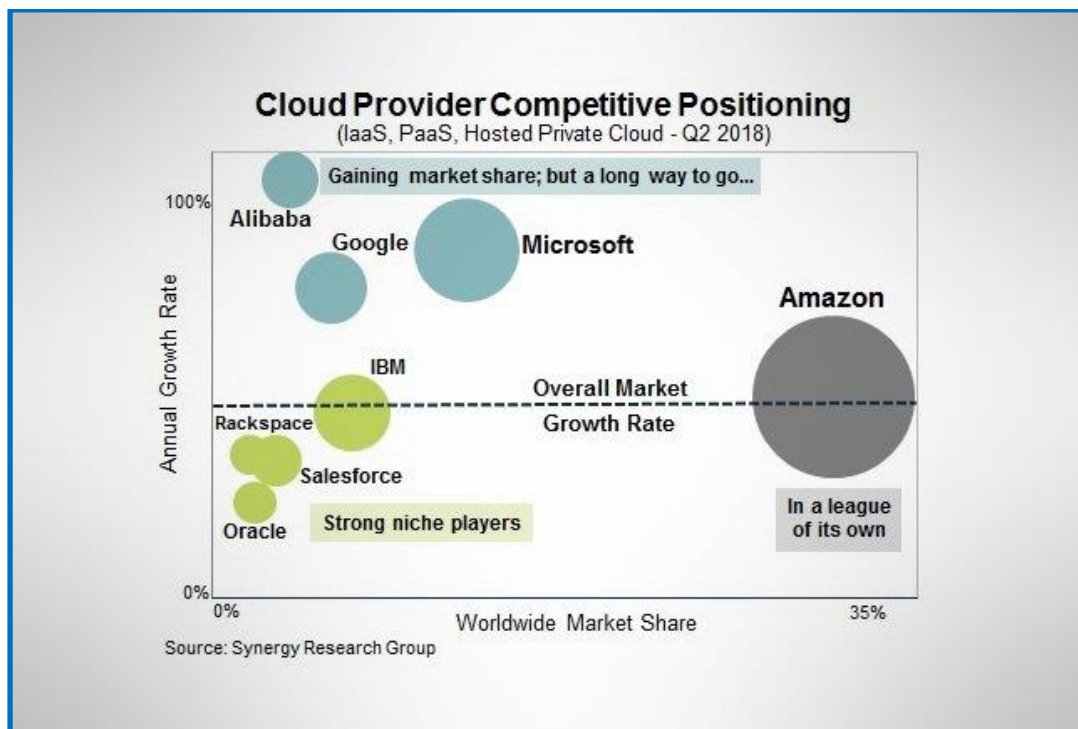
19 : source : <https://fr.outscale.com/>

Il est néanmoins admis que les fournisseurs américains sont les principaux influenceurs des offres cloud publiques actuelles avec aujourd'hui les infrastructures les plus conséquentes à travers le monde et les plus grosses parts de marché.

C'est pourquoi la suite du document sera principalement focalisée sur les offres d'IA « as a service » des trois principaux leaders.



20 : source : Canalys Estimates - Juillet 2018



21 : source : Synergy Research Group - 2018

Cela ne veut pas dire que les offres des concurrents soient moins pertinentes ou moins efficaces, mais l'analyse du catalogue des trois principaux fournisseurs d'offres de cloud public actuel permet de dégager plus facilement les grandes tendances de ce marché.

3.2 LES DIFFERENTES VISIONS DU CLOUD PUBLIC

J'ai demandé à mes interlocuteurs quelles visions ils avaient aujourd'hui des offres cloud, en particulier pour répondre à des besoins d'IA.

J'ai dégagé trois points de vue.

3.2.1 L'accès à des ressources techniques

C'est la vision adoptée la plupart du temps par les utilisateurs finaux des offres clouds. Pour eux, le cloud représente un accès à des ressources physiques, principalement de la puissance de calcul et de l'espace de stockage à volonté. C'est notamment les offres « IaaS » qui se détachent sur cette vision.

Plus que l'accès à de simples machines virtuelles, le cloud public permet l'accès à de nouveaux types de périphériques notamment les GPU et les TPU.

Les GPU pour « Graphique *Processor** Unit » présentent une architecture optimisée pour des calculs simples, mais réalisables sur des multiples données à la fois. Il permet des traitements massivement parallèles qui peuvent correspondre à certaines caractéristiques d'algorithmes de « Machine Learning ». Mais cela reste un usage détourné des GPU, prévu à l'origine pour la gestion de l'écran et des graphismes sur un ordinateur.

On voit donc arriver sur le marché des TPU (Tensor Processor Unit) ou NPU (Neural Processor Unit) dont l'architecture a été spécialement pensée pour la mise en application de réseaux neuronaux. Ces composants sont aujourd'hui relativement chers et pas toujours simples à acquérir pour les entreprises. Le cloud public, comme nous le verrons dans les offres disponibles, propose aujourd'hui l'accès à ces équipements.

3.2.2 Un modèle économique

C'est la vision adoptée principalement par les fournisseurs de cloud public. Le modèle économique proposée permettant une consommation « à la demande » des besoins « IT » (Information technology) a été étendu à l'IA.

Les fournisseurs mettent en avant la possibilité pour tout à chacun d'expérimenter l'IA selon différentes offres sans nécessité d'investissements massifs. J'ai pu obtenir dans mes témoignages des retours sur le **l'importance d'expérimenter l'IA avant d'en conclure à sa nécessité pour ses usages.**

Beaucoup d'entreprises n'ont pas encore d'idée très précise sur ce qu'elles pourraient attendre d'une solution d'IA. Comme nous l'avons vu dans l'historique de l'IA, l'une des causes de l'hiver de l'IA a été la déception des résultats obtenus par rapport aux investissements réalisés. La possibilité d'exploiter des ressources sans engagement pour évaluer une solution d'IA est un encouragement économique fort pour pousser les entreprises à se lancer dans l'IA

3.2.3 Une approche collaborative

C'est une vision que je n'avais pas imaginée, mais qui est revenue plusieurs fois dans mes interviews. Tous mes interlocuteurs s'accordent pour dire que **nous ne sommes qu'au début de l'usage de l'IA**.

Pour que les solutions d'IA évoluent, **il est nécessaire d'expérimenter le plus de cas d'usages possibles et de partager les retours obtenus**. Le cloud public présentant un accès naturellement ouvert vers l'extérieur, il facilite la collaboration entre intervenants.

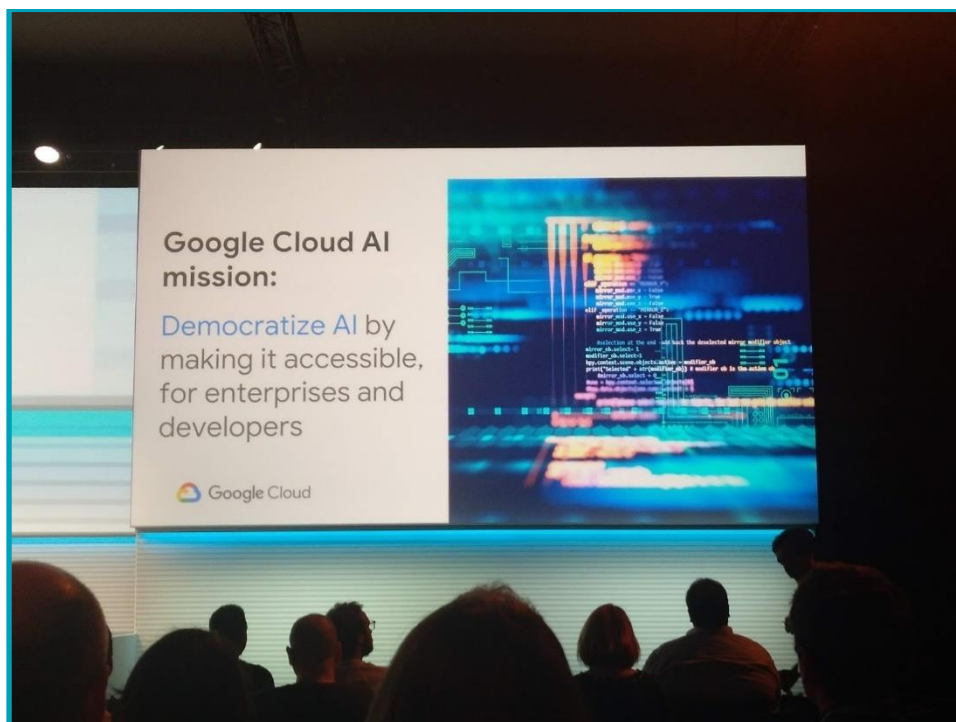
Pour les fournisseurs, de cloud public, comme de solution d'IA, le fait de mutualiser certaines de leurs solutions à destination de plusieurs clients leur permet de faire évoluer leurs produits beaucoup plus vite en capitalisant sur les retours obtenus et les données engrangées.

3.3 LES OFFRES D'IA DANS LE CLOUD

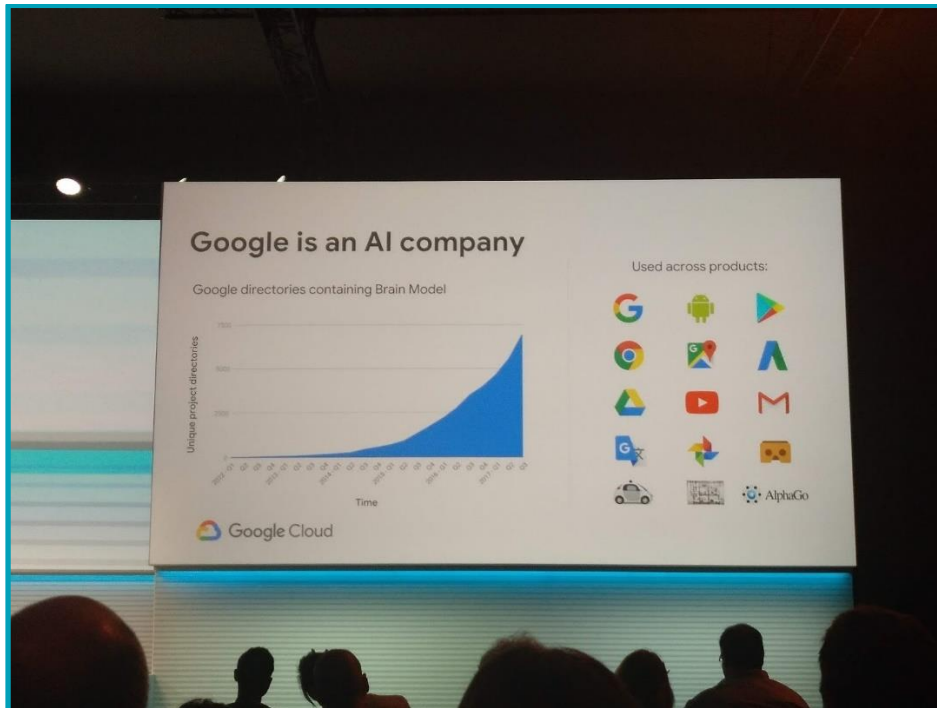
Il est incontestable que les principaux providers de cloud public communiquent énormément autour de leurs solutions dédiées à la mise en œuvre d'IA pour les entreprises.

J'ai pu assister à plusieurs évènements organisés par ces derniers et l'IA apparaît très régulièrement dans leur ligne de produits et dans leurs annonces auprès des professionnels.

Par exemple, Google a présenté beaucoup d'éléments liés à l'IA lors du « Google Cloud Summit » organisé à Paris le 5 juin 2018



22: Photo prise lors du Google Cloud Summit le 5 juin 2018 à paris



23: Photo prise lors du Google Cloud Summit le 5 juin 2018 à Paris

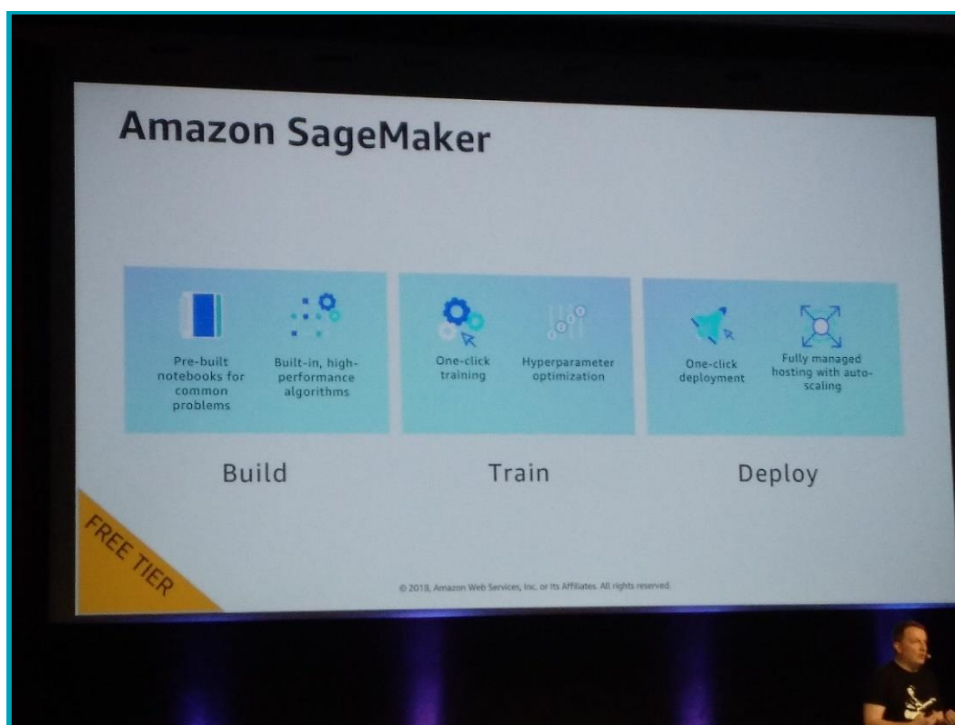


24: Photo prise lors du Google Cloud Summit le 5 juin 2018 à Paris

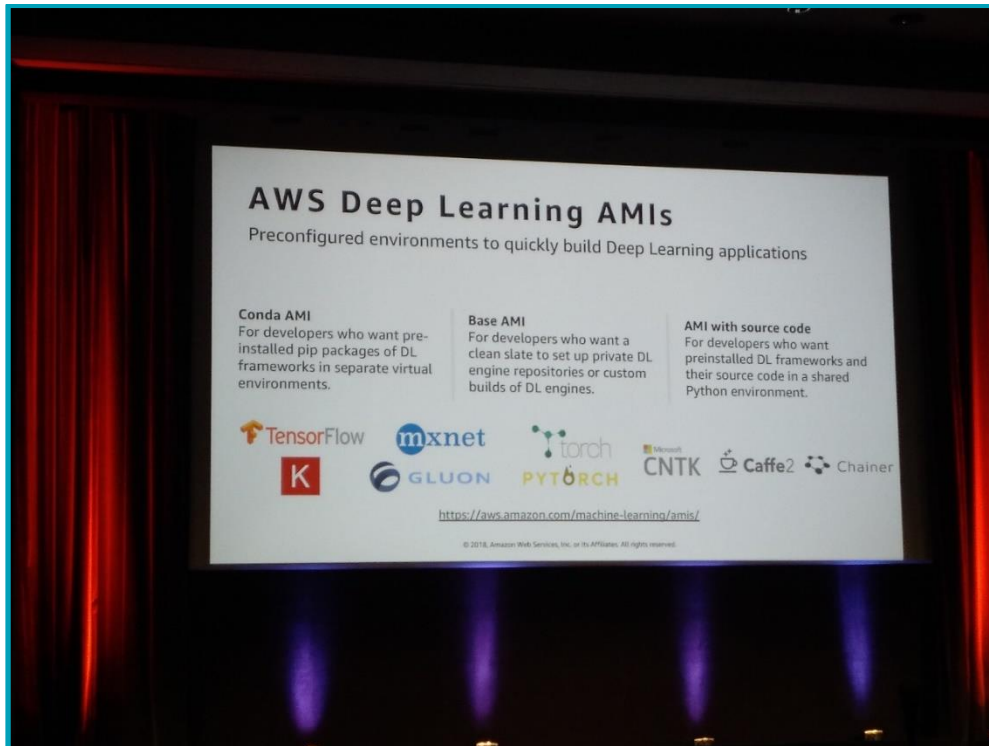
J'ai observé ce même engouement pour l'IA de la part d'Amazon Web Service lors de l' « AWS Summit » 2018 organisée à Paris le 19 Juin



25 : Photo prise lors de l'AWS Summit 2018 à Paris le 19 juin

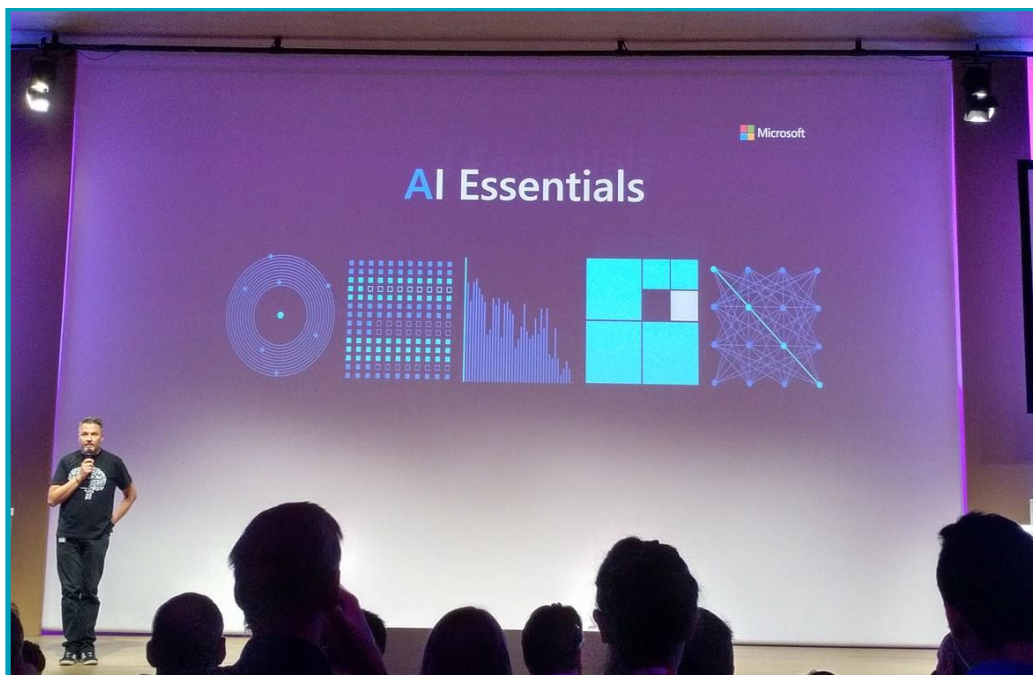


26 : Photo prise lors de l'AWS Summit 2018 à Paris le 19 juin

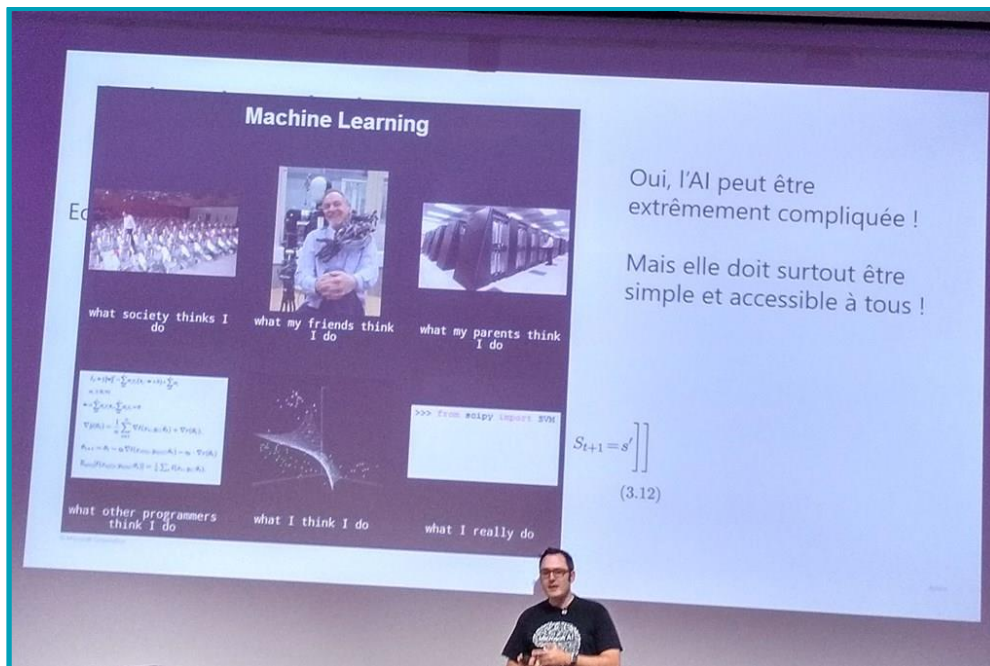


27 : Photo prise lors de l'AWS Summit 2018 à Paris le 19 juin

Microsoft organise depuis peu des évènements spécifiques autour de l'IA appelés « AI* Essentials ». J'ai pu participer à l'un d'eux le 26 septembre 2018 ou j'ai pu constater que le fournisseur du cloud « Azure » s'implique également fortement sur les sujets d'IA.



28 : photo prise lors de l'évènement "AI Essentials" organisé par Microsoft le 26/09/2018 à Paris



29 : photo prise lors de l'évènement "AI Essentials" organisé par Microsoft le 26/09/2018 à Paris

D'une façon générale, **on observe que l'IA est devenue pour les fournisseurs de cloud public un nouveau marché afin d'amener les entreprises à basculer dans leur écosystème.** De la même façon qu'ils se présentaient, il y'a quelques années comme des facilitateurs du « *BigData** », ils se revendiquent maintenant comme des accélérateurs de l'adoption de l'IA dans les entreprises.

J'ai échangé avec mes interlocuteurs durant mes interviews pour connaître leur ressenti sur ces offres et tenter de savoir si oui ou non elles remplissaient bien les promesses affichées dans leurs descriptions.

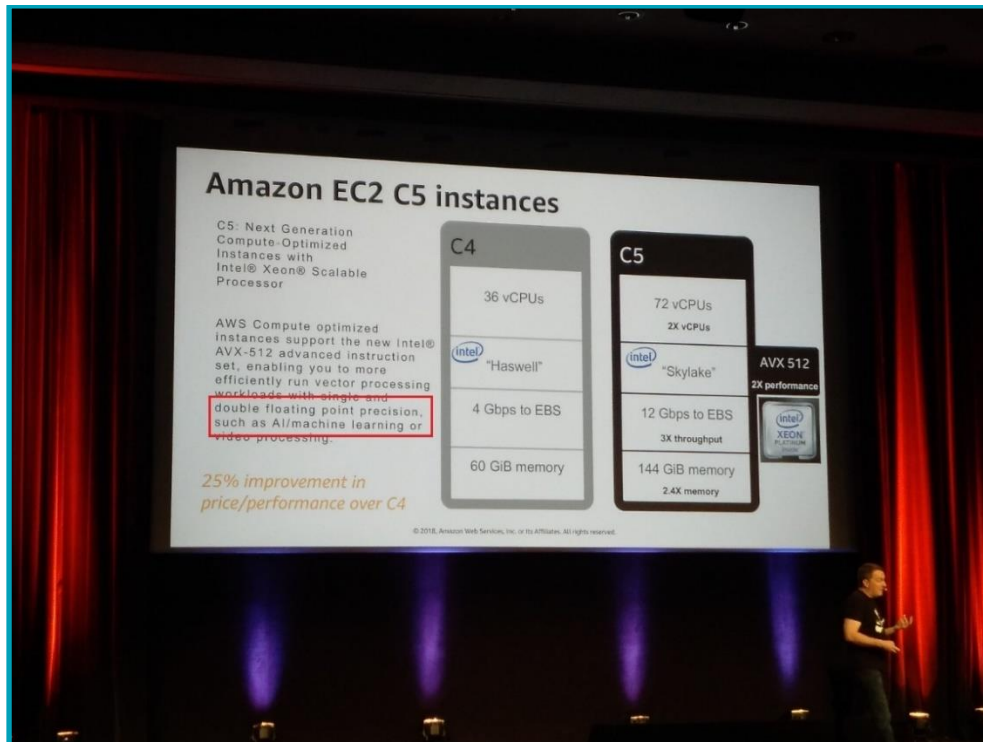
J'en suis arrivé à dégager trois types de solutions, répondant à différents besoins.

3.3.1 Les offres pour clients experts sur l'IA

Ces offres sont celles qui m'ont paru aujourd'hui comme les plus déployées et parallèlement ce sont les plus « basiques » en rapport aux autres propositions du cloud public pour l'IA.

Elles regroupent principalement des solutions IaaS (Infrastructures As A Service) et s'adressent à des entreprises disposant d'un savoir-faire autour de l'IA et d'une expérience avancée du traitement de la donnée. On y retrouve la mise à disposition d'environnements d'exécutions optimisés pour les modèles de « Machines Learning ».

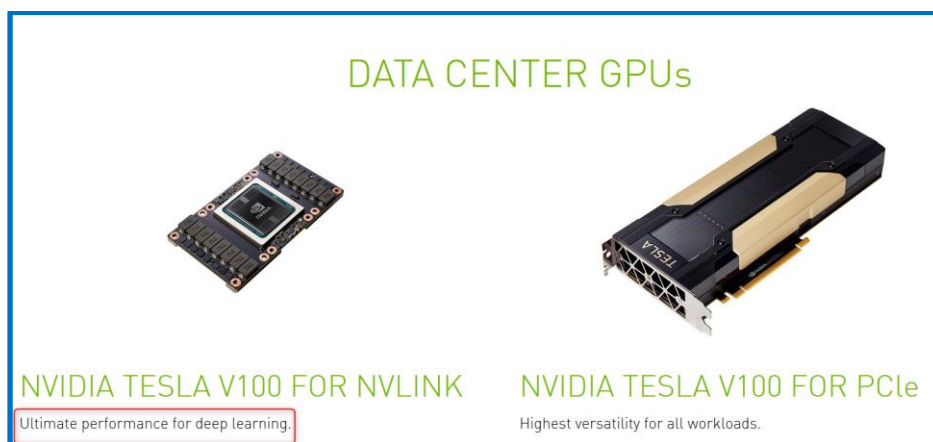
Il s'agit le plus souvent d'instances de VMs* déployées sur des générations d'équipements adaptées aux algorithmes de « Machines Learning »



30 : Exemple d'instance AWS équipée de la dernière génération de processeur Intel

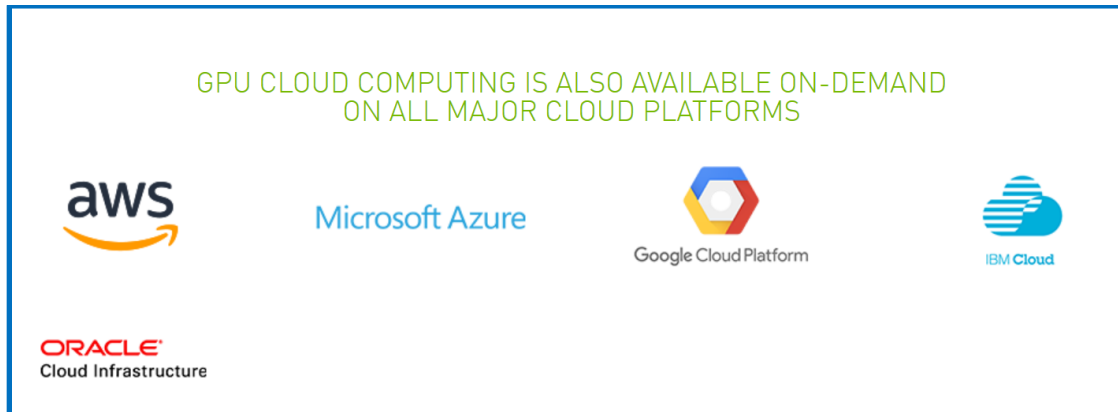
Comme évoqué précédemment, les « GPU » sont venus compléter le catalogue des modèles de VMs des principaux providers de cloud.

L'architecture de ces puces auparavant dédiées aux calculs 3D a trouvé une nouvelle utilisation à travers ses capacités de traitements massivement parallèles particulièrement adaptées à certains algorithmes de « Machine Learning ». Le constructeur « NVIDIA » est un partenaire privilégié des fournisseurs de cloud publics. Leurs cartes graphiques sont reconnues sur le marché, mais les modèles spécifiques aux calculs applicatifs sont souvent chers et nécessitent des serveurs adaptés.



31: source : <https://www.nvidia.com/en-us/data-center/tesla-v100/>

La majorité des providers de cloud public se proposent de mettre à disposition ces équipements afin d'éviter aux entreprises d'en faire l'acquisition



32 : source : <https://www.nvidia.com/en-us/data-center/gpu-cloud-computing/>

Ces offres se retrouvent chez beaucoup de fournisseurs et pas seulement parmi les leaders Américains.

Performance graphique, calcul intensif et intelligence artificielle grâce aux GPU NVIDIA sur le Cloud Outscale

Home > Les offres de services de Cloud Computing > GPU NVIDIA

GPU NVIDIA

GPU NVIDIA pour les architectures hautes performances et l'intelligence artificielle

En plus de l'agilité et de la flexibilité apportées par son Cloud, Outscale met à votre disposition les **GPU (Graphics Processing Unit) NVIDIA**. Des processeurs spécifiquement destinés à fournir des calculs intensifs pour supporter les applications disposant de données volumineuses et complexes.

Les **GPU NVIDIA** fournissent des capacités de calcul exceptionnelles pour traiter différents types de charges de travail (créations graphiques, transactions financières, calculs scientifiques,...). Les GPU NVIDIA sont directement intégrés aux Machines Virtuelles disponibles sur le Cloud Outscale pour booster le traitement de données et accélérer les calculs.

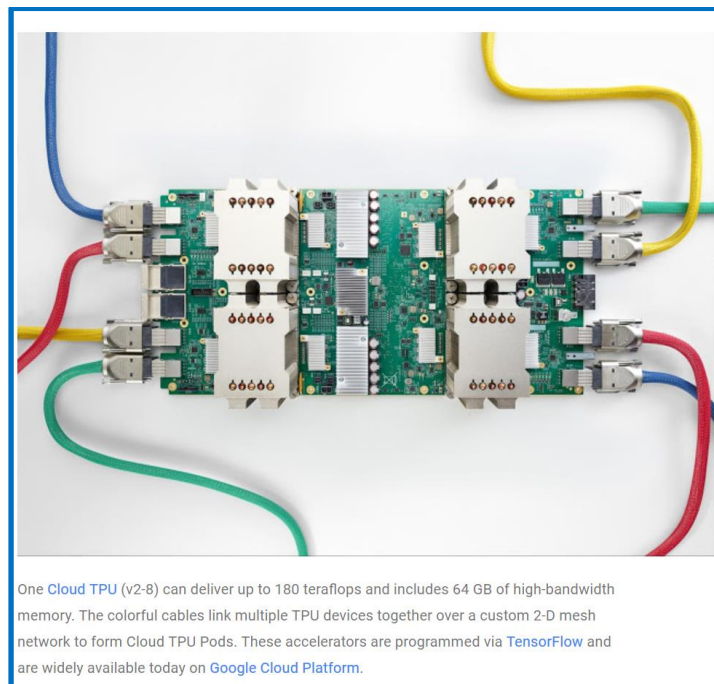
Intégrés dans le Cloud Outscale, les GPU NVIDIA sont capables, grâce aux centaines de milliers de coeurs intégrés, de traiter les informations très rapidement et vous projeter dans des expériences innovantes avec des performances exceptionnelles, pour des usages de **calcul intensif** et de **l'intelligence artificielle**.

En parallèle des **CPU** existants sur les Machines Virtuelles, les processeurs **GPU NVIDIA Grid** s'intègrent dans le Cloud Outscale et rendent ainsi vos applications plus rapides et plus performantes.

33 : Offre GPU chez le fournisseur français "Outscale" : source : <https://fr.outscale.com/les-offres-de-services-de-cloud-computing/gpu/>

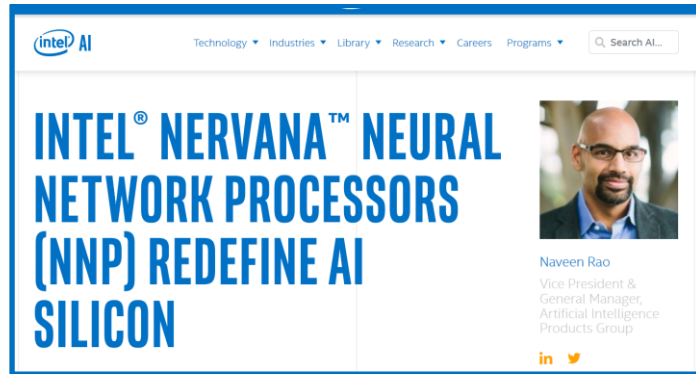
Les GPU ne sont pas les seuls composants disponibles pour la mise en œuvre d'IA. Des processeurs spécialisés pour le traitement de réseaux neuronaux sont désormais disponibles. Ces derniers restent encore assez confidentiels et leur acquisition reste très complexe pour les entreprises.

Google est à l'initiative d'une des puces les plus utilisées dans ce domaine pour l'instant à travers leur « TPU » (Tensor Processing Unit). Elle est spécifiquement adaptée aux programmes d'AI basés sur le *framework** spécialisé « Tensor Flow », également proposé par Google, mais en *open source**. L'usage de cette puce est pour l'instant une exclusivité du cloud de Google.



34: source : <https://cloudplatform.googleblog.com/2018/06/Cloud-TPU-now-offers-preemptible-pricing-and-global-availability.html>

Ce type de processeur va être amené à se développer de plus en plus. Le fabricant « Intel » s'est également positionné sur ce marché via leur futur produit « Nervana »



35 : source : <https://ai.intel.com/intel-nervana-neural-network-processors-nnp-redefine-ai-silicon/>

J'ai appris l'existence de cette puce lors de mon interview avec « Julien Simon, Technical Evangelist Artificial Intelligence » chez AWS. En effet, le numéro 1 du cloud public compte bien pouvoir proposer ce type de processeurs dès leur mise sur le marché à travers son offre IaaS.

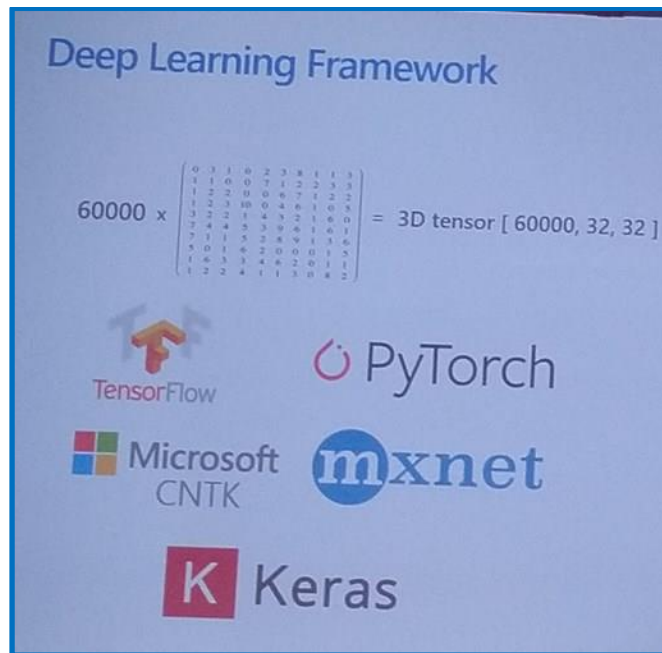
Il n'y a pas que sur le matériel que les clouds providers cherchent à proposer des solutions optimisées pour l'IA. On trouve aujourd'hui de nombreux *frameworks** (ensemble d'outil et de fonctions) dédiés à la programmation de « Machine Learning », comme « TensorFlow » pour faciliter le travail des développeurs. Certaines instances de *VM** (virtual machine) proposées par les clouds *providers**, en plus d'être équipées de CPU modernes et de puces spécialisées (GPU ou TPU) sont disponibles avec des versions de système d'exploitation intégrant les *frameworks** de développement dans des versions spécifiquement optimisées par le *provider** pour l'exécution du code dans son écosystème « IaaS ».

Choosing an AWS Deep Learning AMI

Even for experienced machine learning practitioners, getting started with deep learning can be time consuming and cumbersome. The AMIs we offer support the various needs of developers. To help guide you through the getting started process, also visit the [AMI selection guide](#) and more [deep learning resources](#).

<h3>CONDA AMI</h3> <p>For developers who want pre-installed pip packages of deep learning frameworks in separate virtual environments, the Conda-based AMI is available in Ubuntu, Amazon Linux and Windows 2016 versions.</p> <p>Learn more about the benefits of the Conda AMI and get started with this step-by-step guide.</p>	<h3>BASE AMI</h3> <p>For developers who want a clean slate to set up private deep learning engine repositories or custom builds of deep learning engines, the Base AMI is available in Ubuntu and Amazon Linux versions.</p> <p>Learn more about the benefits of the Base AMI and get started with this step-by-step guide.</p>
--	---

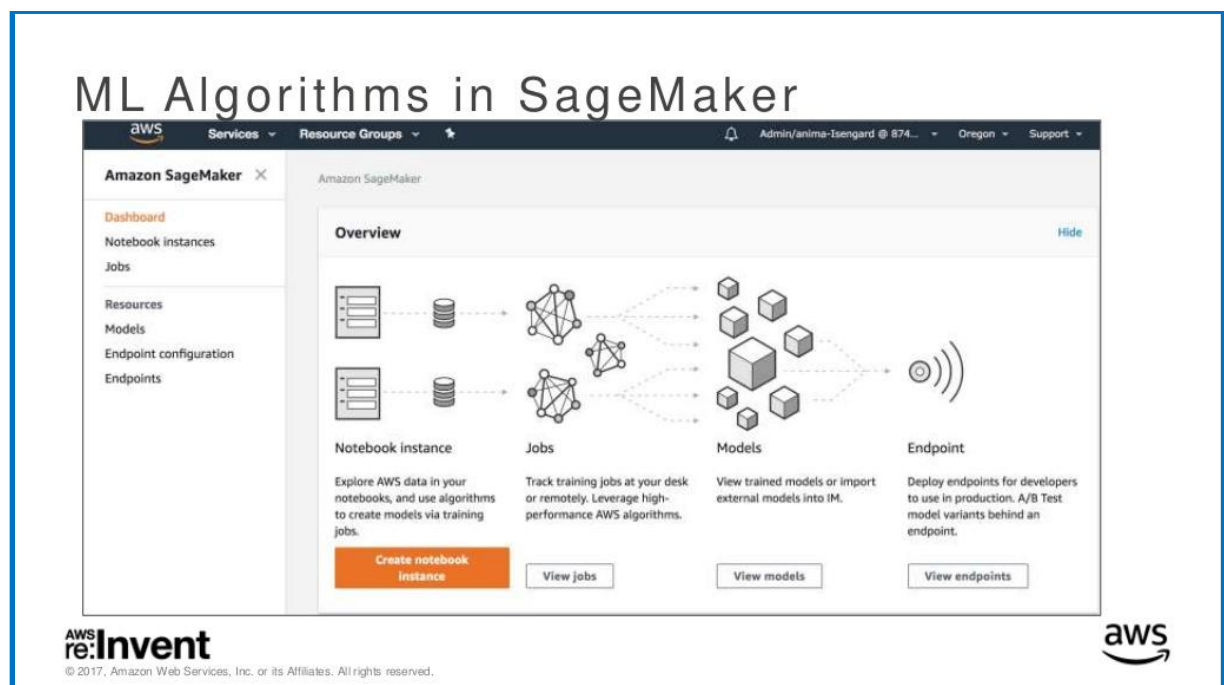
36 : source : <https://aws.amazon.com/machine-learning/amis/> (AMI = Amazon Machine Image : image d'OS maintenue et proposée par AWS)



37 : Framework pris en charge dans Microsoft Azure : source : conférence "AI Essentials"

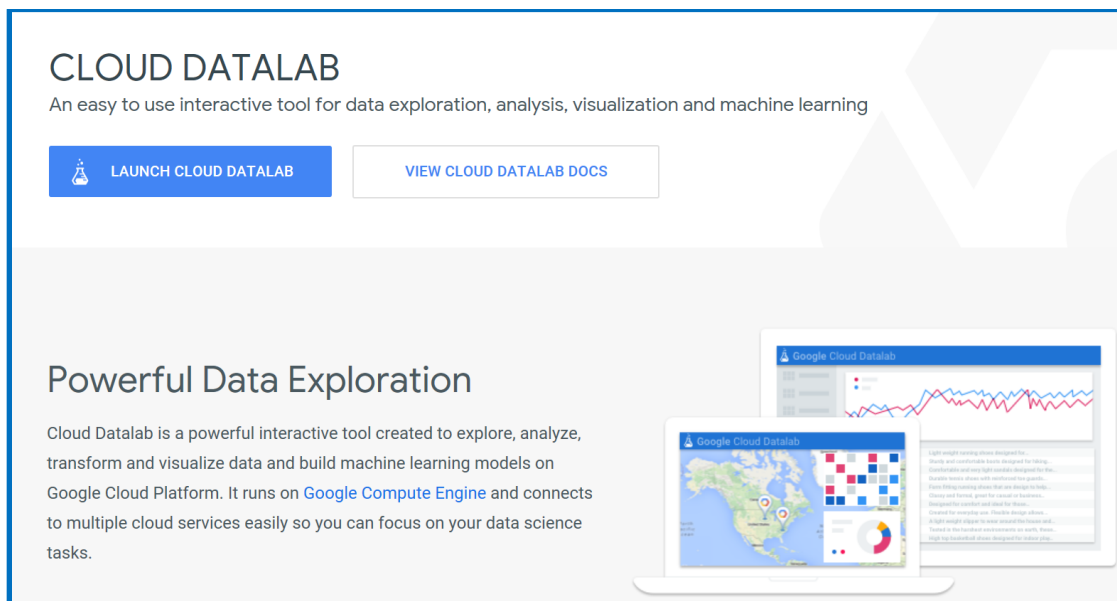
Ces solutions mises à dispositions des entreprises par les cloud providers sont le plus souvent secondés par des environnements de développement spécifiquement conçus pour faciliter le codage de ses propres modèles de machines learning

C'est par exemple le cas de « AWS Sage Maker »



38 : source : AWS : <https://aws.amazon.com/fr/sagemaker/>

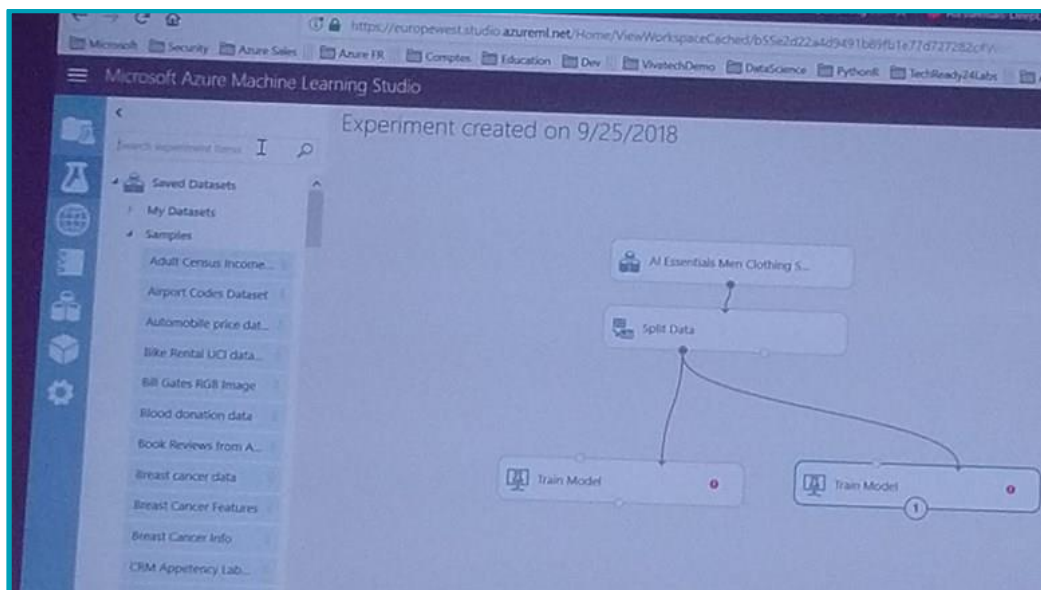
Le cloud Google propose sa solution « Cloud DataLab »



39 : source : <https://cloud.google.com/datalab/>

Ces plateformes facilitent le cycle de vie des modèles algorithmiques développés pour la mise en œuvre d'AI. C'est une réponse des clouds providers à la difficulté exprimée par de nombreuses entreprises ayant débuté le développement d'une solution d'IA pour le passage en production de leurs modèles et le suivi de leur cycle de vie.

Elles permettent également la construction d'algorithmes d'IA via des interfaces simplifiées par exemple la solution « Azure Machine Learning » de Microsoft.



40 : Usage de "Azure Machine Learning" : source : conférence "AI Essentials"

En résumé, **les clouds providers sont à l'écoute des nouveautés matérielles optimisées pour l'IA et s'efforcent de les mettre à disposition de leurs clients.**

Pour faciliter leurs exploitations et leurs usages, les clouds providers proposent une couche additionnelle dédiée aux développements et aux mises en productions. On retrouve donc les avantages du IaaS, à savoir le provisionnement à la demande d'instance optimisée pour l'IA, secondé par une offre « PaaS » (plateforme As A Service).

Ces offres permettent aux entreprises **disposant des compétences nécessaires** de créer, tester et déployer leur solution d'IA sans avoir à investir dans une plateforme matérielle spécifique.

Les clouds provider "packagent" et optimisent des solutions logiciels « *Open Sources** » pour offrir aux développeurs des écosystèmes connus et réputés. Néanmoins, ces offres ne peuvent pas convenir aux sociétés qui ne disposent pas de compétences internes autour du développement d'algorithme de machine Learning.

Malgré l'avantage des outils proposés par les clouds providers, il reste nécessaire pour les entreprises qui souhaitent exploiter ces offres d'avoir des personnes disposant d'un savoir-faire fort autour de l'IA et du traitement de la data, comme des *data scientist** ou des développeurs spécialisés. S'ajoute à cela que les instances IaaS associées à ces offres **sont souvent parmi les plus chères proposées aux catalogues** des providers.

Tout le monde n'a pas forcément besoin d'un tel niveau de puissance et comme nous l'avons vu précédemment certains modèles de machines learning ne nécessitent pas d'équipements spécialisés et peuvent convenir à des cas d'usage rencontrés en entreprises. Certains *data scientists** que j'ai rencontrés arrivent à proposer des modèles tournant sur de simples PC portables.

Penser qu'il est obligatoire de disposer d'équipement ultra performant de dernière génération pour proposer une solution d'IA est une erreur. Tout dépend des cas d'usages, des volumétries de données à traiter et du type d'algorithme déployé. Un réseau neuronal n'est pas forcément nécessaire et certaines entreprises peuvent exploiter des solutions d'IA plus basiques, mais qui peuvent répondre parfaitement à leurs usages. L'un des risques de ces offres serait de céder aux arguments parfois un peu exagérés des fournisseurs de clouds, vantant la simplicité d'usage de leur écosystème pouvant amener finalement à des coûts importants sans résultat si le client n'a pas les compétences nécessaires. Pour les entreprises suffisamment matures ou souhaitant simplement expérimenter quelques cas d'usage d'IA, ces offres sont par contre extrêmement intéressantes et **accélèrent grandement le développement de modèles de « Machine Learning »**

Je vais maintenant traiter d'un autre type d'offres cette fois-ci plus simple d'approche en termes d'exploitation et d'usage.

3.3.2 Les offres « clef en main »

L'écriture applicative d'une IA à partir de zéro n'est pas simple. Beaucoup d'entreprises peuvent prétendre à des besoins d'IA sans pour autant avoir les ressources humaines nécessaires à sa mise en œuvre, même pour des cas de figure basiques.

Certains besoins sont répétitifs, comme la traduction de texte par exemple. Il est alors plus simple de se tourner vers des offres spécialisées sur le sujet, dont beaucoup sont portées par les principaux acteurs du cloud public.

En effet, Google, Amazon, Azure et d'autres fournissent des accès à des API spécialisées dans le traitement d'une problématique pouvant bénéficier d'une IA. (Une API, pour « **A**pplication **P**rogramming **I**nterface » permet d'interagir de manière simple avec une application pour lui soumettre une demande et récupérer un résultat)

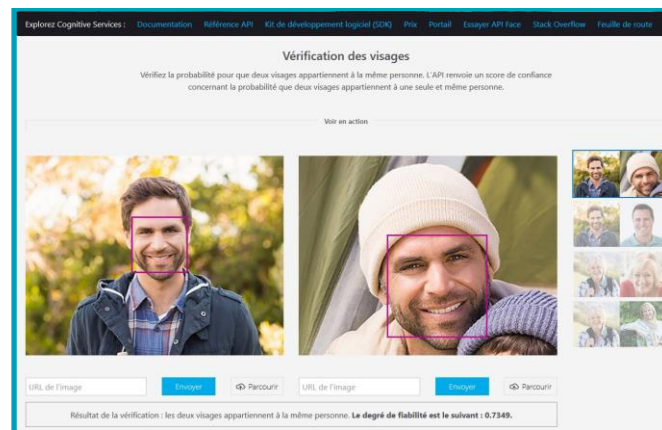
Dans ce cas de figure, le client n'a qu'à exploiter ces API dans ses applications sans avoir à coder le moindre élément d'IA ou algorithme de machine Learning.

Lorsqu'il appelle l'API, ce sont les modèles d'IA des clouds providers qui sont sollicités et qui renvoient les résultats attendus.

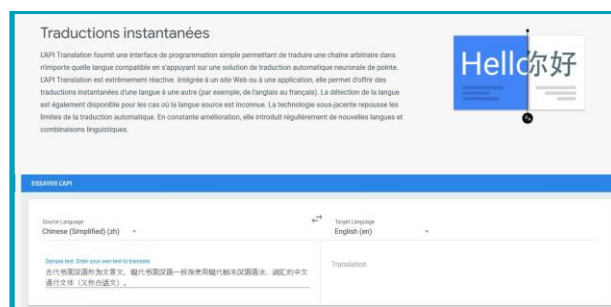
J'ai ainsi pu assister à des démonstrations impressionnantes chez AWS où

- 1) L'application codée par l'entreprise prend une photo d'un panneau publicitaire contenant du texte
- 2) L'image est envoyée chez le cloud provider via l'API de ce dernier dédiée à la reconnaissance d'image. Cette dernière est traitée dans le cloud et l'IA associée extrait le texte de l'image.
- 3) Le texte de l'image est reçu par l'application de l'entreprise. Il est réémis dans cloud, mais cette fois-ci via l'API du provider dédié à la traduction de texte.
- 4) L'application de l'entreprise réceptionne la traduction du texte dans 6 langues différentes et renvoie un des résultats obtenus à nouveau dans le cloud à destination de l'API en charge du « Text To Speech ».
- 5) L'application de l'entreprise exploite le résultat de l'appel à la dernière API pour énoncer à haute voix le texte de l'image dans la langue choisie

Au final, l'entreprise n'aura eu à coder que de simples appels aux différentes API et l'ensemble du code de l'application reste basique et facilement maintenable. Toute l'intelligence nécessaire à la reconnaissance d'image, la traduction et l'énonciation du texte est laissée à la responsabilité du cloud provider. Celui-ci refacture l'entreprise en fonction du nombre d'appels à ses API*.



41 : exemple de service de traitement d'image chez Azure: source : <https://azure.microsoft.com/fr-fr/services/cognitive-services/face/>



42 : Exemple d'API de traduction disponible chez Google : source : <https://cloud.google.com/translate/>

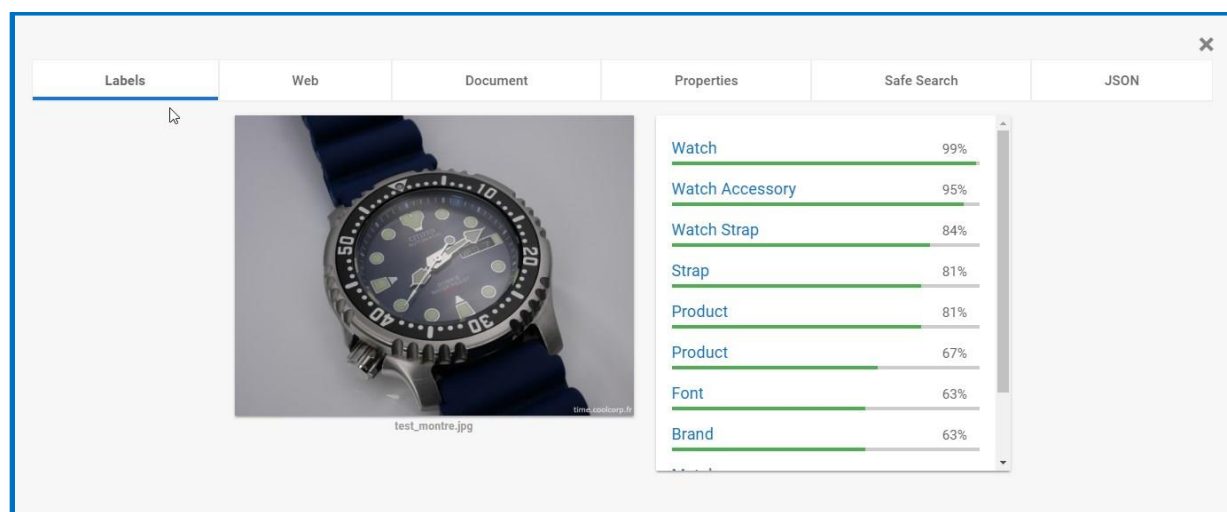
En dépit de la simplicité d'exécution de ces offres, **la plupart de mes interlocuteurs se sont montrés méfiants sur l'usage de ces services**. Il m'a souvent été rappelé qu'une IA efficace est une IA qui s'est construite avec un modèle de données de qualité et conforme aux vocabulaires de l'entreprise qui l'exploite.

Trois défauts remontent sur l'usage de ces API* « généralistes »

- **L'usage de données génériques comme moteur d'apprentissage**. Les API permettent d'accéder à des moteurs d'IA entraînée sur des données mutualisées issues des soumissions des clients et des bases de connaissances du provider. Cela peut poser problème. Par exemple, il paraît évident pour un humain que le mot « livret » n'ait pas forcément le même sens selon qu'il soit employé dans un milieu bancaire, médical ou littéraire. Ces possibles différences de sens sont beaucoup plus complexes à appréhender pour une IA. L'usage de service d'« IA sur étagère » ne peut répondre aux problématiques contextuelles des entreprises qui réclament des versions spécifiques des modèles de machine Learning déployés.
- Un système de facturation pouvant évoluer. **L'application cliente côté entreprise ne dispose d'aucune intelligence dans son code**. Elle est dépendante à 100% du service cloud et si ce dernier est amené à disparaître ou à adopter un nouveau modèle tarifaire, il se peut que l'application cliente devienne inopérante ou trop chère à exploiter.
- L'usage de ses données. Lorsque vous exploitez ce type d'offres, vos données vont participer à l'alimentation des modèles algorithmiques mis en œuvre par le cloud provider. Vous participez donc indirectement à l'amélioration d'un service pour lequel vous êtes vous-même facturé. Bien que des clauses contractuelles existent et que certaines offres peuvent vous garantir la non-participation de vos données à l'amélioration de l'outil du provider, c'est un sujet auquel il faut réfléchir.

En résumé, les offres d'« IA » « clef en main » des différents providers de clouds sont **extrêmement intéressantes pour un usage rapide et générique de quelques cas d'usage de l'IA**. Elles permettent d'ajouter des capacités d'IA à des applications existantes ou nouvelles sans nécessiter de compétences fortes dans les entreprises. Elles ne peuvent néanmoins être exploitées dans des cas de figure où l'IA est un enjeu majeur pour une société et qui désire obtenir un modèle spécifique à son activité. C'est un point que j'ai pu confirmer auprès des *fintech** spécialisées que j'ai interviewées. Si elles exploitent principalement une architecture cloud publique pour fonctionner, leurs moteurs d'IA n'exploitent pas les API proposées et tournent sur des offres IaaS de plus bas niveau.

En complément, pour illustrer ces limitations, j'ai simplement testé l'API* de reconnaissance d'image de Google avec la photo d'une montre.



43 : Exemple de résultat obtenu avec l'API de reconnaissance de Google

L'API nous retourne la probabilité de 99% que l'image analysée soit bien une montre... Si ce résultat est correct et peut se montrer déjà intéressant, on n'a aucune information complémentaire qui permettrait de différencier cette montre d'une autre. Pas de reconnaissance de la marque, du modèle et on s'aperçoit assez vite de la limite de la solution.

3.3.3 Les offres assistées

Les deux types d'offres précédentes présentent soit l'inconvénient de nécessiter de l'expertise interne soit de ne pas être parfaitement adaptées aux natures des données à traiter.

Les cloud provider travaillent donc à de nouvelles offres qui pourraient être simples et rapides à opérer tout en s'adaptant aux spécificités contextuelles de chaque entreprise.

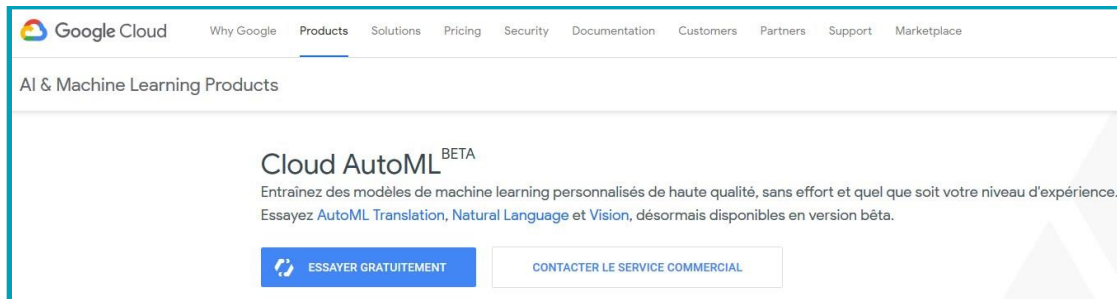
Le principe de ces offres est le suivant :

- 1) L'entreprise fournit ses propres données
- 2) Grâce à un ensemble d'outils simples et accessibles, l'entreprise commence à entraîner son propre modèle de « Machine Learning » sur son jeu de données
- 3) Une fois le modèle opérationnel, l'entreprise peut le valider sur un nouveau jeu de données qui lui serait propre et décider ou non de mettre ce modèle en production à travers une ou plusieurs applications.

On retrouve ainsi la simplicité des offres « clef en main » avec la possibilité d'appeler un modèle de « Machine Learning » à travers une simple API. Mais ce modèle n'est pas générique et exploité par plusieurs clients, il est unique à l'entreprise.

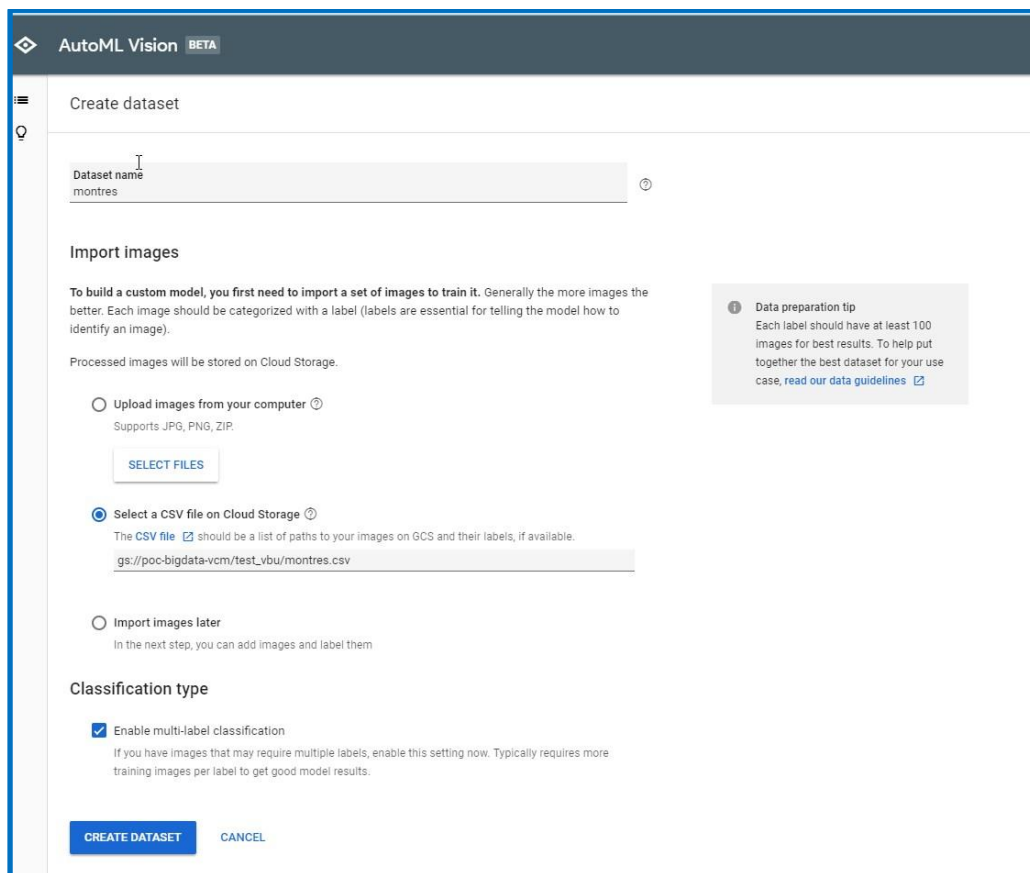
Il repose sur un apprentissage entièrement dépendant des données de l'entreprise, et peut donc se montrer plus efficace et compatible avec le domaine d'activité spécifique de cette dernière.

Ces offres sont pour le moment assez limitées et disponibles le plus souvent en version bêta. Google propose par exemple sa solution nommée AutoML.



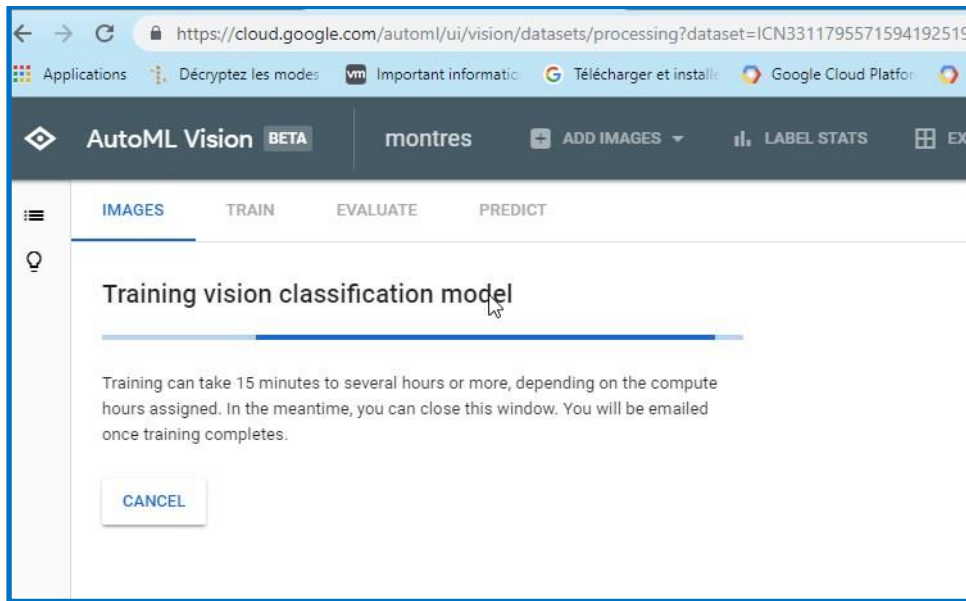
44 : source : <https://cloud.google.com/automl/>

À titre de comparaison, j'ai repris la même photo de montre que celle utilisée pour le test de l'API de reconnaissance d'image, mais je l'ai soumise cette fois-ci à la solution AutoML de Google. Pour cela j'ai d'abord fourni un référentiel de 800 photos de montres que j'ai labélisé par marque, modèles et type de montre (montre de plongée, montre classique...)



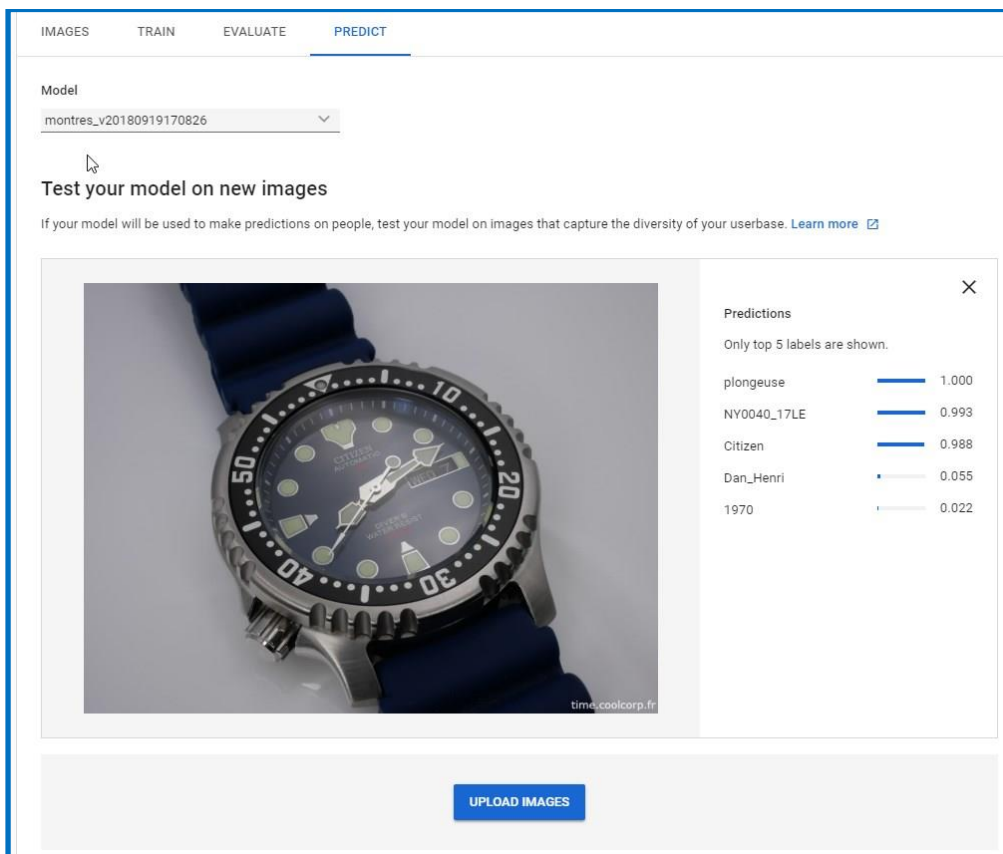
45 : Écran d'importation des échantillons d'images de AutoML

J'ai pu ensuite lancer l'entraînement du modèle sur mon échantillon de photos.



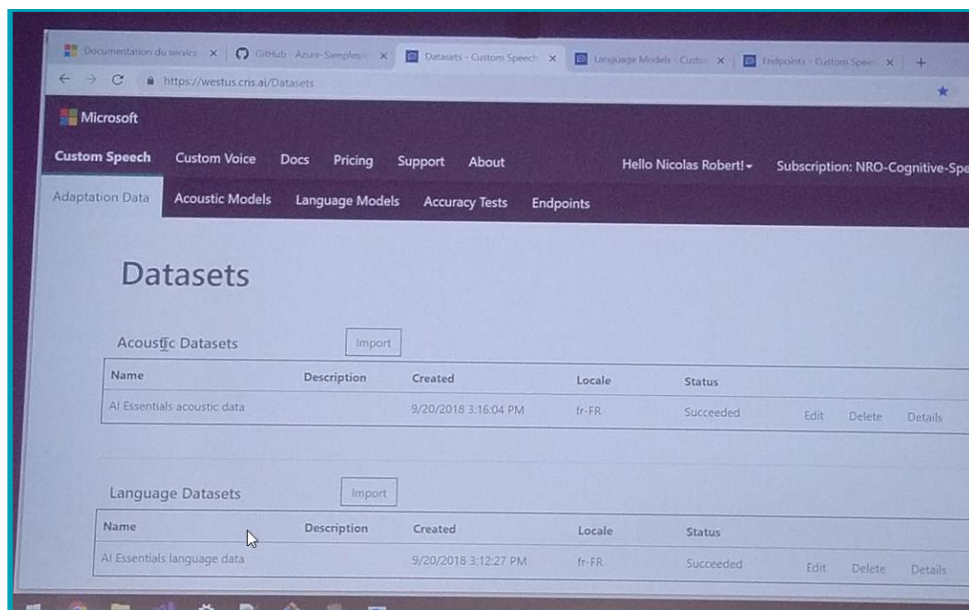
46 : Entraînement du modèle dans AutoML

J'ai finalement testé mon modèle sur la photo témoin.



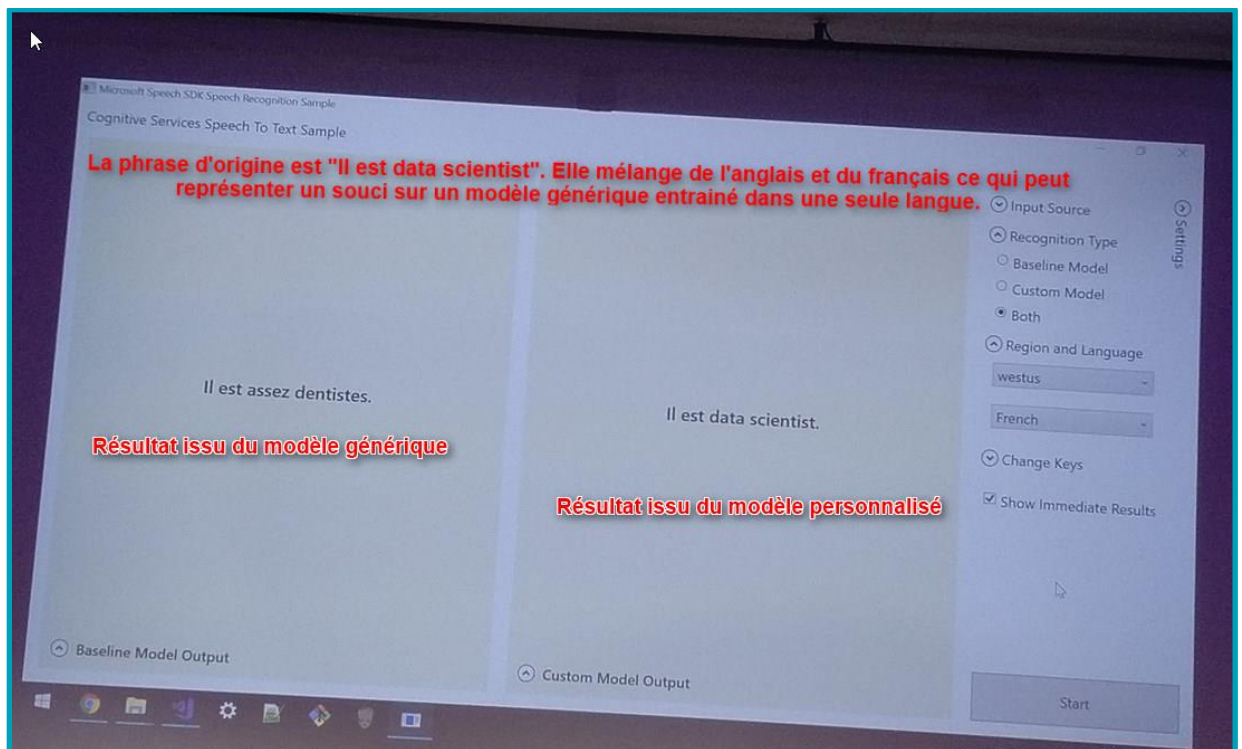
47 : résultat dans AutoML

Il est par exemple possible d'exploiter l'API de reconnaissance de la parole générique disponible dans Microsoft Azure puis de lui appliquer des échantillons supplémentaires pour le rendre compatible avec un vocabulaire personnalisé



49 : Service "Custom Speech" de Microsoft Azure : source: conférence "AI Essentials"

Cela va permettre d'exploiter un modèle unique interrogeable par ses propres applications.



50 : Application du modèle personnalisé : source conférence "AI Essentials"

Malgré l'efficacité apparente et la grande facilité d'exploitation de ce type de solutions, elles possèdent néanmoins certains défauts :

- **Le modèle de « Machine Learning » est uniquement exécutable au sein du cloud public du provider.** Vos applications ne peuvent que le solliciter via l'API associée et le code fourni par la solution. Vous devenez donc dépendant à 100% de l'écosystème du fournisseur pour faire fonctionner votre IA. La réversibilité est très complexe, voire impossible.
- **Vous n'avez pas la connaissance des critères utilisés par le modèle de « Machine Learning » pour classifier vos données.** Ce n'est pas un défaut propre à ce type d'offres, mais une limitation des algorithmes de Machine Learning utilisés. Il semble bien que le deep Learning soit employé par les fournisseurs de cloud (à minima pour la reconnaissance d'image) et l'on retrouve les soucis d'explicabilité évoqués précédemment dans ce document. Il est alors possible que l'« IA » exploite des critères totalement étrangers à l'analyse souhaitée. Par exemple, si l'on cherche à construire une IA capable d'apprendre à partir de photos de chiens à identifier les différentes races automatiquement, il est tout fait possible que les algorithmes mis en œuvre tiennent compte du cadre des images, du lieu et des conditions de prises de vues pour identifier les animaux. Si les échantillons d'images de « Huski » montrent tous l'animal dans la neige, cette dernière peut alors devenir un critère déterminant pour identifier un "Huski".
- **La simplicité d'utilisation n'implique pas une simplification des données nécessaires.** Pour que ces modèles de « Machine Learning assistés » soient efficaces, il est primordial d'avoir un maximum d'échantillons et que ces derniers soient correctement classifiés et variés. Dans mon exemple de photos de montres, dès que je soumettais à mon modèle un cliché d'une montre avec un cadran bleu, j'obtenais toujours en sortie une indication proche de 100% de la détection d'une montre de plongée. Les deux seules montres de plongée présentes dans mon échantillon d'apprentissage possédant un cadran bleu, j'en ai déduit que cette caractéristique est devenue pour l'« IA » une méthode de reconnaissance d'une montre de plongée.
- **Ces solutions sont pour l'instant encore « jeunes » et limitées à des cas d'usage spécifiques.** On manque encore de retour et d'expérience pour juger de l'efficacité des modèles obtenues une fois en production.

Il est indéniable que ce type d'offres présente une efficacité de mise en œuvre d'une solution d'IA, même lorsque les technologies sous-jacentes exploitent du Deep Learning.

Elles masquent totalement la complexité des modèles algorithmiques et les contraintes d'architectures techniques normalement nécessaires pour la mise en œuvre de ce type d'IA.

Il reste néanmoins nécessaire pour le client d'avoir une **excellente connaissance de ses données et d'avoir les compétences nécessaires ainsi qu'un corpus de données d'apprentissage suffisant pour tirer parti de ces offres.**

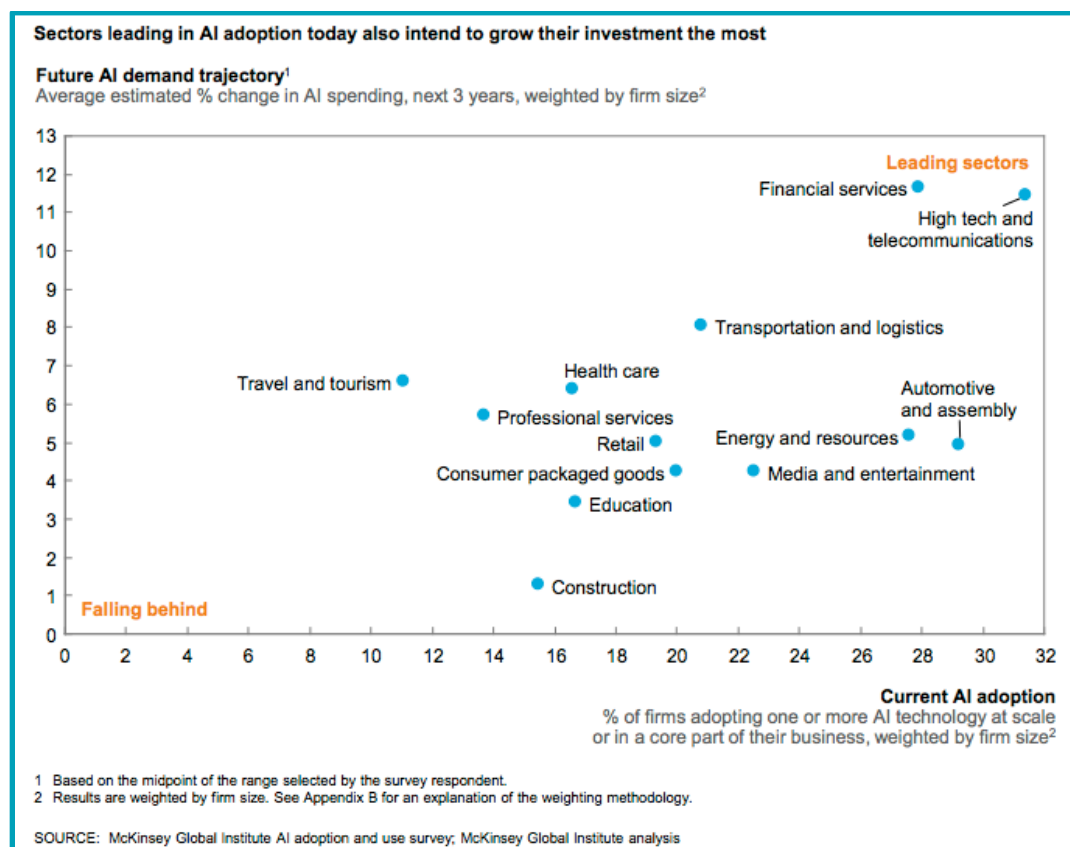
Il doit être également conscient que si l'AI est pour lui un facteur de développement capital, utiliser ce type de solutions lui impose un lien très fort avec le provider qu'il aura choisi. Même l'usage des API génériques permet une réversibilité plus « simple » en cas de désaccord avec son fournisseur.

Cela revient aux mêmes problématiques que celles associées aux services managés proposés par les principaux fournisseurs de cloud. Plus vous exploitez des services spécifiques à un fournisseur vous rapprochant d'un modèle "SaaS", plus vous devenez dépendant de votre fournisseur. Ce n'est pas forcément un problème, mais **c'est un point qu'il est important d'analyser avant toute souscription de ce type d'offres.**

4 IA et CLOUD, une projection possible chez un Asset Manager

Comme constaté, le cloud public offre un ensemble d'offres orientées vers la mise en œuvre d'IA. Il serait maintenant intéressant de se projeter dans un contexte professionnel pour imaginer comment pourraient s'intégrer ces offres et quels avantages elles apporteraient à l'entreprise.

J'ai choisi pour cela de traiter le cas d'une société d'Asset Management. Son activité s'inscrit dans le secteur de la finance, un secteur particulièrement intéressé par l'IA.



51 : Document du 31 aout 2017 : source : McKinsey Global Institute

Il s'agit d'imaginer des cas d'usages possibles : je n'ai pas mis en application les exemples qui vont suivre. Il faut rester prudent quant à l'efficacité réelle des propositions. C'est néanmoins un exercice intéressant pour juger ou non de la pertinence des offres des clouds providers dans la mise en application d'IA.

4.1 LE MÉTIER D'ASSET MANAGER

Une entreprise d'« Asset Management » ou de Gestion de portefeuilles (aussi appelée Gestion d'actifs) est une société chargée de la gestion financière administrative et comptable de produits de placements pour des tiers. Les particuliers et les entreprises peuvent lui confier de l'argent pour qu'il soit géré à travers différents fonds. Ses clients lui confient ce qu'on appelle un mandat qui lui permet de prendre des décisions de placements selon les termes d'un contrat. Ses décisions s'appliquent sur un portefeuille représentant le capital du client. Elle intervient sur un marché réglementé et s'engage à gérer de manière indépendante et dans l'intérêt exclusif de l'investisseur les sommes qui lui sont confiées.

Son organisation peut varier d'une société à une autre. Elle peut se spécialiser sur certains types de marchés financiers ou s'adresser à une clientèle spécifique.

On retrouve néanmoins chez tout « Asset manager » deux axes d'activités.

- **La gestion de l'actif** : c'est le cœur de métier qui regroupe l'ensemble des activités liées à la gestion financière (sélections des investissements, achat/vente d'actifs...).
- **La gestion du passif** : c'est le regroupement des activités liées à la centralisation des ordres d'investissement et/ou des opérations de règlements

Ces activités sont réparties à travers trois principales fonctions

- Le front Office : c'est la fonction de gestion. On y retrouve le métier de « gérant » chargé de prendre les décisions d'investissement.
- Le Middle Office : c'est la fonction en charge du traitement administratif des transactions négociées par le front office
- Le back office : c'est la fonction en charge de l'administration des mandats, c'est-à-dire du calcul de la valeur des portefeuilles

Ces métiers s'appuient sur des fonctions de supports comme le marketing, le contrôle interne et autres.



Source : AFG.

52 : source: AFG

Comme on peut le constater, une société d'Asset Management est composée de plusieurs métiers, intervenant à différents niveaux de la chaîne de valeur. Je vais me concentrer dans les parties suivantes sur trois métiers en particulier dans lequel j'ai essayé de trouver des exemples de mises en application d'IA à travers des offres cloud (ce qui n'exclut pas d'autres cas d'applications possibles sur d'autres métiers)

4.2 UN CHATBOT AU SERVICE CLIENT

4.2.1 Le chatbot

Un chatbot ou agent conversationnel est un programme en charge de dialoguer avec une personne humaine. Ce n'est pas un concept moderne, puisque le premier chatbot date de 1966 comme on a pu l'observer dans l'historique de l'IA. Néanmoins les chatbots ont connu un regain d'intérêt important ces dernières années. En effet, si au début les chatbots répondaient sans chercher à comprendre le sens de l'information traitée, les évolutions de l'IA actuelle lui permettent de se rapprocher de plus en plus d'une véritable compréhension de l'information.

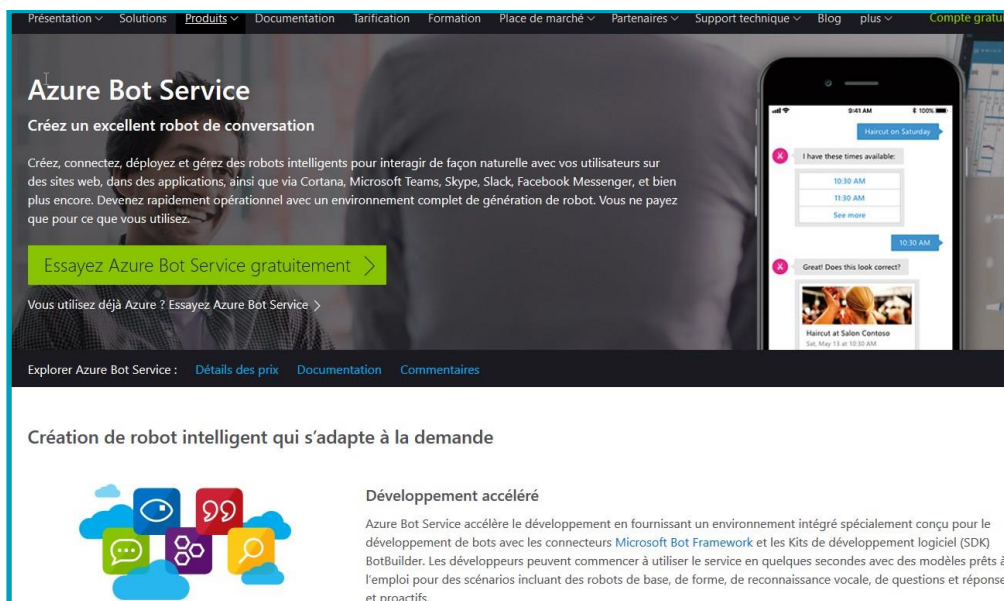
4.2.2 Intérêts

Dans le cas d'une société d'Asset Management, un chatbot pourrait très bien s'intégrer sur le site institutionnel de l'entreprise pour guider le client dans le choix des offres disponibles.

Il pourrait également servir de point d'entrée pour le service client, permettant de traiter les questions les plus courantes de façon automatique pour rediriger les cas les plus complexes vers un interlocuteur humain.

4.2.3 Mise en application dans le cloud public

Le chatbot pourrait être mis en service directement à partir d'un service clé en main chez un cloud provider, par exemple la solution « Azure Bot Service »



Présentation Solutions Produits Documentation Tarification Formation Place de marché Partenaires Support technique Blog plus Compte gratuit

Azure Bot Service

Créez un excellent robot de conversation

Créez, connectez, déployez et gérez des robots intelligents pour interagir de façon naturelle avec vos utilisateurs sur des sites web, dans des applications, ainsi que via Cortana, Microsoft Teams, Skype, Slack, Facebook Messenger, et bien plus encore. Devenez rapidement opérationnel avec un environnement complet de génération de robot. Vous ne payez que pour ce que vous utilisez.

Essayez Azure Bot Service gratuitement >

Vous utilisez déjà Azure ? Essayez Azure Bot Service >

Explorer Azure Bot Service : Détails des prix Documentation Commentaires

Création de robot intelligent qui s'adapte à la demande

Développement accéléré


Azure Bot Service accélère le développement en fournissant un environnement intégré spécialement conçu pour le développement de bots avec les connecteurs Microsoft Bot Framework et les Kits de développement logiciel (SDK) BotBuilder. Les développeurs peuvent commencer à utiliser le service en quelques secondes avec des modèles prêts à l'emploi pour des scénarios incluant des robots de base, de forme, de reconnaissance vocale, de questions et réponses et proactifs.

53 : source : <https://azure.microsoft.com/fr-fr/services/bot-service/>

Ce qui rend la mise en application complexe chez un « Asset Manager », c'est que le secteur de l'entreprise exploite **un vocabulaire bien spécifique**. On retrouve des termes comme « OPCVM* », « VL* » avec des mots qui peuvent prendre un sens particulier comme « portefeuille ».

On peut alors « améliorer » le chatbot pour le rendre davantage compatible avec l'activité de l'entreprise en ajoutant une brique additionnelle du fournisseur nommé « Language Understanding »

Avantages




Ajoutez la compréhension du langage naturel à vos applications

Conçue pour identifier les informations importantes d'une conversation, la Reconnaissance vocale interprète les objectifs de l'utilisateur (intentions) et distille des informations utiles à partir de phrases (entités), afin de produire un modèle linguistique nuancé de haute qualité. Language Understanding s'intègre de façon fluide avec le service Speech pour le traitement instantané de la conversion de sortie orale en intention, ainsi qu'avec Azure Bot Service, ce qui facilite la création d'un bot sophistiqué.

Créez rapidement une solution linguistique personnalisée

De puissants outils de développement sont combinés avec des applications et dictionnaires d'entités pré-générés personnalisables, tels que Calendrier, Musique et Appareils, pour vous permettre de créer et déployer une solution plus rapidement. Les dictionnaires extraits de la connaissance collective du web fournissent des milliards d'entrées, et aident votre modèle à identifier correctement les informations importantes dans les conversations des utilisateurs.



54 : source : <https://azure.microsoft.com/fr-fr/services/cognitive-services/language-understanding-intelligent-service/>

On s'aperçoit alors qu'en combinant des solutions orientées « API » mélangeant des offres « clef en main » et des offres « personnalisables », on pourrait offrir un chatbot adapté au secteur de l'Asset Management.

4.3 UN ASSISTANT A LA PRISE DE DECISION POUR LES GERANTS

4.3.1 Le gérant

Le métier de gérant est au cœur de l'activité d'une société d'« Asset Management ». C'est lui qui décide de la stratégie d'investissement applicable au portefeuille du client en fonction du mandat qui lui a été donné et des contraintes réglementaires.

Pour prendre les bonnes décisions, le gérant est obligé de s'appuyer sur de nombreux indices financiers et se construire des référentiels d'informations. Les données qu'il peut utiliser peuvent être dans le domaine public ou fournies par des sociétés spécialisées

4.3.2 Intérêts.

Les sources de données pouvant aider aux prises de décisions d'un gérant ne cessent de se multiplier et de se diversifier. Si les informations financières ont toujours constitué le référentiel d'un gérant, la concurrence entre « Asset Manager » oblige désormais à considérer d'autres types de sources de données, parfois plus exotiques et non structurées. L'avènement des réseaux sociaux a fait exploser les indices potentiels de tendance de marché, et il n'est plus forcément évident pour un gérant de se reposer uniquement sur les fournisseurs spécialisés pour constituer son parcours de décisions.

Dans ce contexte, l'assistance d'une IA pour la mise en avant du signal d'investissement caché dans les flux d'informations pourrait permettre au gérant de se démarquer de ses concurrents. L'IA pourrait également lui permettre d'étendre son référentiel de données et de considérer de nouveaux critères de choix d'investissement.

4.3.3 Mise en application dans le cloud public

Cet exemple est particulier, car il touche au cœur du savoir-faire d'une société d'Asset Management. Nous sommes typiquement dans le cas où l'IA à mettre en œuvre doit reposer sur **des modèles extrêmement travaillés et conçus au plus proche de l'activité du gérant**. Il y'a également **une notion de propriété intellectuelle**, car les données à traiter sont multiples et une partie de ces dernières est liée au patrimoine de la société d'Asset Management.

Dans ce contexte, on pourrait donc imaginer que l'entreprise aurait tout intérêt à concevoir elle-même son application de machine learning et se reposer sur le cloud public uniquement sur les offres de IaaS. Elle pourrait ainsi investir sur l'humain et acquérir les compétences internes nécessaires à la mise en œuvre de sa propre solution tout en tirant parti des performances du cloud public.

La phase d'apprentissage, qu'on imagine longue dans ce cas, pourrait tirer parti des instances de calcul et des espaces de stockage d'un provider comme AWS.

Par exemple en exploitant des instances de VM* spécialisées disposant de GPU

Instances Amazon EC2 P3

Accélérer l'apprentissage automatique et les applications de calcul haute performance avec des GPU puissants.

Démarez avec les instances P3

De grandes entreprises comme Airbnb, Salesforce et Western Digital utilisent les instances Amazon EC2 P3 pour alimenter leurs applications d'apprentissage automatique et de calcul haute performance.

Les instances Amazon EC2 P3 offrent les meilleures performances de calcul dans le cloud, sont rentables, prennent en charge l'ensemble des principales structures d'apprentissage automatique et sont disponibles dans le monde entier.

Propulsé par jusqu'à huit GPU NVIDIA Tesla V100 de dernière génération, les instances Amazon EC2 P3 délivrent jusqu'à 1 pétaflop de performances de précision mixte par instance pour accélérer significativement les applications d'apprentissage automatique et de calcul haute performance. Il a été prouvé que les instances Amazon EC2 P3 réduisent le temps d'apprentissage automatique de quelques jours à quelques minutes, ainsi que le temps d'obtention des résultats pour des calculs haute performance.

88 % des projets TensorFlow dans le cloud fonctionnent sur AWS.

Dans ce rapport, Nucleus Research donne cinq raisons pour les techniciens de choisir AWS pour l'apprentissage profond plutôt que d'autres prestataires cloud.

Lire le rapport

VISIONNEZ : Apprenez-en davantage sur les instances Amazon EC2 P3 et découvrez comment Airbnb l'utilise pour alimenter ses applications d'apprentissage automatique (48:08 min).

55: source : <https://aws.amazon.com/fr/ec2/instance-types/p3/>

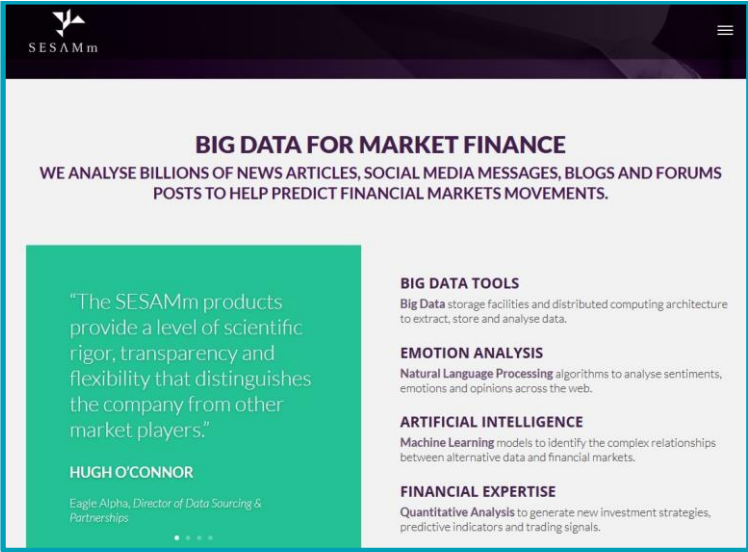
De cette manière, l'IA pourrait être développée plus rapidement pour éventuellement être ensuite appliquée non plus sur le cloud, mais dans l'infrastructure interne de l'entreprise.

La phase d'apprentissage étant terminée avec une totale maîtrise du modèle de ML associé au code déployé, l'entreprise d'Asset Management pourrait appliquer la phase d'exploitation en interne si elle ne souhaite pas laisser l'IA tourner au sein d'un fournisseur externe.

L'inconvénient reste cependant la nécessité d'avoir de très fortes compétences humaines en interne. **L'exploitation de la donnée et le choix d'algorithmes adaptés sont capitaux. Envoyer toutes les sources d'informations sans exploration et sans nettoyage préalable conduirait à un modèle totalement inefficace. Le « Machine Learning » peut être très performant, mais il n'est pas « magique ».** Dans le cas d'une mise en œuvre d'IA partant de zéro, **l'investissement humain reste extrêmement important.**

Pour tenter de gagner en simplicité sans forcément tomber dans les offres cloud clef en main ou personnalisables, la société d'Asset Management pourrait se faire accompagner dans son projet par une fintech spécialisée. Beaucoup de startups arrivent sur le marché pour proposer des moteurs d'IA spécifiquement conçus pour répondre à des problématiques métiers très orientées.

C'est par exemple le cas de « Sesamm » dont vous pourrez retrouver une interview de leur CEO en annexe du document.



56 : source : <https://www.sesamm.com/>

Dans ce cas de figure, c'est toujours le cloud public qui est utilisé, mais de manière indirecte puisque la société d'« Asset Management » passe alors par un partenaire spécialisé qui lui-même exploite des ressources « IaaS » d'un fournisseur de cloud



57 : source : <https://www.mindfintech.fr/article/841/sylvain-forte-sesamm-technologie-danalyse-demotions-permet-aux-banques-dinvestissement-de-predire-mouvements-marches/>



58 : source : <http://paperjam.lu/communiqu/sesamm-va-utiliser-le-cloud-ibm>

4.4 LE SUIVI DU PARCOURS CLIENT AU MARKETING

4.4.1 Comprendre le raisonnement du client

Le dernier exemple est plus généraliste et pourrait s'appliquer à de nombreux secteurs. Il s'agit de mieux comprendre le comportement de ses clients. On souhaite exploiter les informations dont on dispose sur ces derniers pour y appliquer des modèles de machine learning. On agrège tout ce qu'il est possible d'agréger comme donnée sur ses clients : statistiques de navigation du site web de l'entreprise, fiches de contact, historique d'investissement, informations contractuelles, commentaires...**en s'assurant de leur qualité et de leur véracité !**

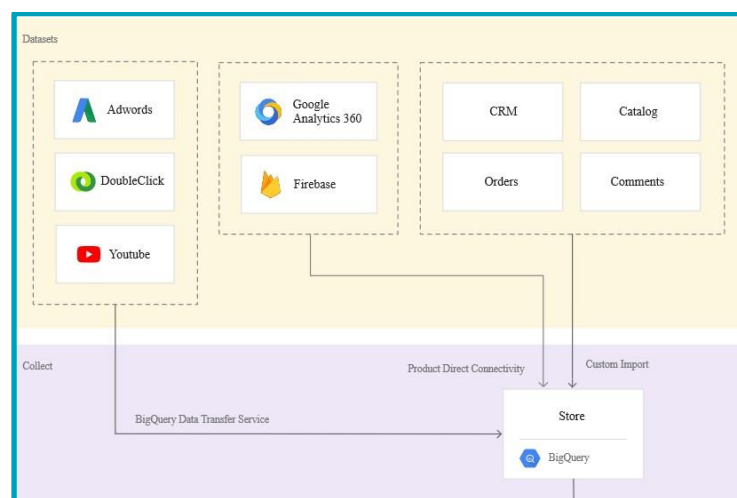
4.4.2 Intérêts

Mieux connaître ses clients et tenter de prédire leurs comportements permet d'établir des offres optimisées et personnalisées. L'IA peut mettre en avant des facteurs jusque-là inconnus du marketing pour améliorer sa stratégie commerciale. Dans le cas de l'Asset management, c'est par exemple définir de nouveaux fonds d'investissements présentant des statistiques importantes de réussite sur le marché ou à même de séduire une clientèle spécifique.

4.4.3 Mise en application dans le cloud public

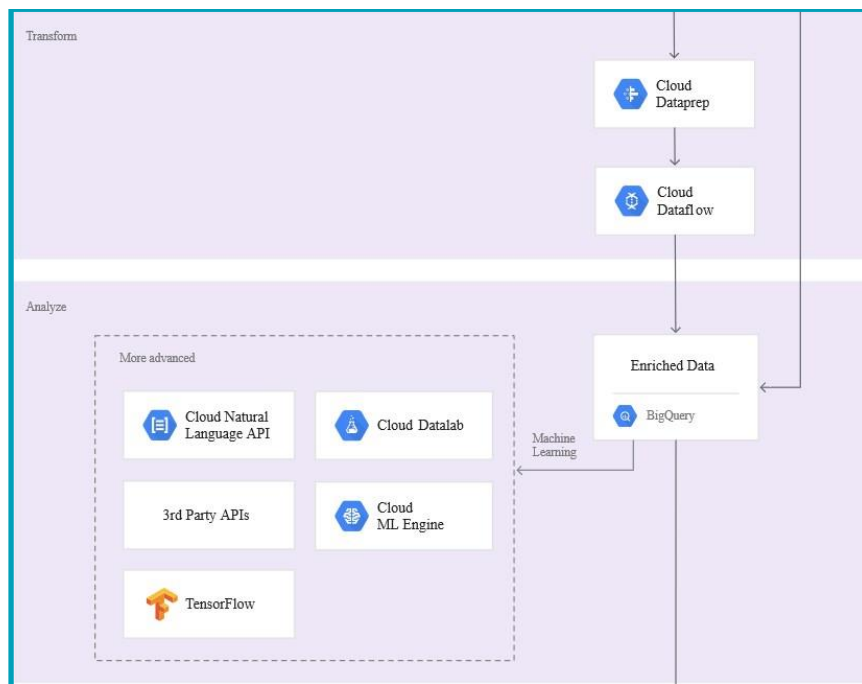
Si l'on souhaite prédire au mieux le comportement de ses clients et identifier des facteurs différenciant, il faut maximiser ses sources de données en s'assurant de leur qualité. Introduire trop d'éléments incomplets, erronés peut mettre à mal l'IA finale. Il y'a donc un travail important d'identifications des sources d'informations et de préparations de la donnée en amont de la phase d'apprentissage de son modèle de Machine Learning.

Pour cela, le cloud public dispose déjà d'un grand nombre d'outils que j'ai évoqué indirectement lorsque j'ai parlé du BigData. Des produits comme « BigQuery » de Google permettent la mise en œuvre d'entrepôt de données en mode SaaS sans se préoccuper des contraintes de stockage et de performance.



59 : source : <https://cloud.google.com/solutions/marketing-data-warehouse-on-gcp>

Ce qui est avantageux, c'est qu'une fois ses données dans l'écosystème du provider, on peut lui appliquer les technologies d'analyses orientées machine Learning intégrées aux outils de manipulation de la donnée.



60 : source : <https://cloud.google.com/solutions/marketing-data-warehouse-on-gcp>

L'entreprise d'Asset Management n'a pas à revoir son architecture d'entrepôt de données pour lui ajouter des capacités de « Machine Learning ». Elle capitalise sur les solutions bigdata du fournisseur de cloud qui se voit compléter des offres de « Machines Learning » du même provider.

BigQuery ML : du ML directement dans BigQuery et avec SQL

Google permet aux utilisateurs de BigQuery de développer des modèles de Machine Learning avec SQL, directement depuis l'entrepôt de données cloud via une technologie BigQuery ML



Craig Stedman
Senior Executive Editor
30 jui. 2018

Suivez-nous:

[Twitter](#)
[G+](#)
[LinkedIn](#)
[Email](#)

Quand SQL rattrape le ML. A l'occasion de sa conférence Google Cloud Next, Google a publié une version bêta de BigQuery ML, un service dont la vocation est de permettre de bâtir des modèles de [Machine Learning](#) (ML) depuis BigQuery, mais en ayant recours à des commandes SQL standard.

Actualités

Open Compute : AT&T va publier les spécifications de ses routeurs/passereles 5G

Violin Systems se relance dans le stockage flash à haute performance avec ses baies XV 58

61 : source : <https://www.lemagit.fr/actualites/252445807/BigQuery-ML-du-ML-directement-dans-BigQuery-et-avec-SQL>

On s'aperçoit alors que les trois types d'offres dédiées à l'IA que j'ai traités dans la partie précédente peuvent se retrouver en compléments d'offres existantes et mis en œuvre par certains clients pour le traitement de leurs données.

5 Conclusion

La première partie de ce document m'a permis de proposer une définition moderne de l'IA rattachée aux principales technologies que l'on cherche à déployer aujourd'hui, principalement autour du "Machine Learning".

Elle a mis en évidence le lien fort entre l'IA et la donnée et définie les besoins actuels pour le déploiement de l'IA en entreprise.

Ces besoins ont été mis en adéquation avec l'analyse des différentes offres des principaux fournisseurs de cloud public dans la deuxième partie.

Afin d'illustrer mes propos, la troisième partie a traité de cas d'usages possibles basés sur certaines des solutions cloud publiques vues précédemment. L'ensemble m'a permis de dégager les conclusions suivantes.

5.1 DES OFFRES EN DEVENIR ET UN IAAS TOUJOURS PRESENT

Les offres du cloud public destinées à l'IA suivent la même évolution que le cloud public lui-même. Les fournisseurs ont d'abord intégré des offres « IaaS » afin de construire (ou améliorer) une infrastructure optimisée et performante à même d'absorber les besoins induits par les algorithmes d'IA les plus contraignants (usage des réseaux de neurones).

Le IaaS semble être aujourd'hui le premier vecteur d'attrait des entreprises pour l'exploitation du cloud public dans des besoins d'IA. Avec la montée en puissance des GPU et autres puces spécialisées, le IaaS continue d'attirer des clients vers les fournisseurs de cloud public.

Ces offres ont très vite été rejointes par des solutions de plus haut niveau avec la mise à disposition de plateforme « PaaS » à même de faciliter le travail des développeurs et des datascientist, avec une très forte orientation vers l'open source.

Nous en sommes maintenant à l'apparition d'offres encore plus managées, qu'on pourrait presque qualifier de « SaaS » tant elles tendent à offrir aux entreprises un accès « clef en main » à l'IA.

On arrive désormais à des produits « IA As A Service », mais qui viennent se confronter aux problématiques du « Machine Learning », à savoir le besoin d'avoir des modèles de plus en plus personnalisés et construits sur les données des clients et non plus sur des échantillons génériques.

Les fournisseurs de cloud l'ont bien compris et chacun essaye désormais d'associer simplicité et personnalisation. Il y'a donc de fortes chances de voir continuer les principaux fournisseurs de clouds se disputer le marché via des solutions d'IA de plus en plus managées et assistées. Les offres IaaS ne disparaîtront pas pour autant et seront toujours nécessaires pour répondre aux besoins de certains clients qui souhaiteront toujours garder la main sur leur IA et bénéficier avant tout de la puissance brute de calcul et du stockage du cloud.

5.2 CLOUD ET IA : UNE ASSOCIATION QUI FAIT MOUCHE, MAIS AVEC LAQUELLE IL FAUT RESTER PRUDENT

5.2.1 Un environnement optimisé

À la question « Est -il obligatoire d'exploiter le cloud public pour faire de l'IA ? » la réponse est **NON**. Mais à la question « Est-il intéressant d'exploiter le cloud public pour faire de l'IA ? », la réponse est **OUI**.

Nous sommes actuellement au début de l'IA moderne. Peu d'entreprises aujourd'hui peuvent se prétendre suffisamment matures sur le sujet. **L'expérimentation est essentielle** pour permettre à chacun de se familiariser avec le "Machine Learning" et s'essayer à l'IA. Le cloud public est parfait pour cela, car il propose des offres capables de s'adresser à plusieurs publics, débutant comme expert. **Sa capacité de traitement est sans égale et permet de se focaliser sur les cas d'usage plutôt que sur la mise en place des solutions techniques.**

Il est indéniable que le cloud public a participé à l'essor de l'IA telle qu'on la connaît aujourd'hui, ne serait-ce que par l'investissement que réalisent les principaux fournisseurs de cloud public dans l'IA. AWS, Google et Microsoft n'ont cessé d'acquérir des sociétés spécialisées, participant ainsi aux développements autour de l'IA.

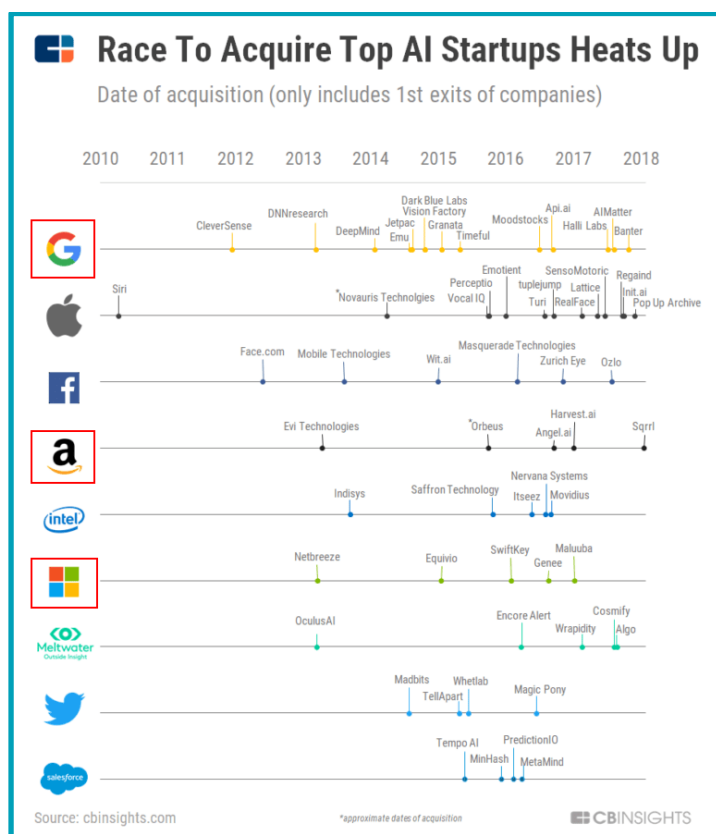


Figure 62 : Historiques des acquisitions de startup spécialisées dans l'IA : source : cbinsights.com

Ce n'est certes pas un composant indispensable pour la mise en œuvre d'IA, mais c'est un moyen d'accélérer sa conception et sa mise en œuvre.

5.2.2 Une simplicité parfois dangereuse

Cette facilité d'accès qu'offre le cloud public à l'IA demeure néanmoins dangereuse. **Il y'a un risque d'imaginer que le simple fait de souscrire une offre cloud spécialisée "Machine Learning" suffit à la mise en œuvre d'une IA et à son succès en entreprise.**

Le cloud public reste avant tout un outil qu'il est possible d'exploiter de mille façons différentes. S'engager dans des offres qui proposent des coûts d'entrée, certes minimes par rapport à des solutions plus traditionnelles, peut parfois impliquer une sensation d'obligation de résultats.

Fort des discours marketing et des exemples d'IA révolutionnaire « propulsés » sur internet, on peut s'emprisonner dans des expérimentations en boucle et finir par engloutir des budgets conséquents pour finalement n'arriver à rien.

J'ai pu m'apercevoir lors de mes interviews des problématiques que pouvaient ressentir certains fournisseurs de solution d'IA face à des clients totalement dans l'illusion des capacités réelles de l'IA aujourd'hui et surtout sur ce qu'elle pourrait leur rapporter.

Même les principaux providers de cloud public, pourtant rarement avares d'arguments marketing, s'emploient de plus en plus à informer des vraies possibilités du Machine Learning et surtout des efforts à faire en amont de son projet d'IA, notamment sur sa donnée.

L'IA a déjà connu plusieurs « hivers » ou de trop grandes promesses ont fini par la transformer en mirage. Plus personne ne souhaite la voir à nouveau s'éteindre au vu des capacités techniques dont on dispose désormais.

En résumé, l'IA facilite l'innovation, mais n'innove pas à la place des entreprises, elle fonctionne sur des projets mûris.

5.2.3 Le côté obscur du cloud public

Il est également important de rappeler que l'usage du cloud public n'est pas toujours anodin. Dans un marché principalement porté par des entreprises américaines, suivi de près par des géants chinois, **il convient parfois de se poser quelques questions autour du besoin de confidentialité de ses données et de ses contraintes de réversibilité.**

Il n'est pas nécessaire de verser dans la « paranoïa », mais dans un secteur comme l'IA ou la donnée est capitale, il est important d'étudier les offres **pas uniquement sur leurs capacités technologiques, mais également sur leurs caractéristiques juridiques.**

5.3 LA FORMATION ET LES COMPETENCES HUMAINES : LES COMPLEMENTS PARFAITS A L'IA DANS LE CLOUD PUBLIC

S'informer des réalités de l'IA, partager ses expériences et acquérir des compétences restent essentiels à une bonne mise en œuvre de l'IA en entreprise.

Dans ce cas, le cloud public est un formidable moyen de construire sa solution d'IA ou à minima de trouver un partenaire capable de répondre à ses besoins. Il n'est pas forcément nécessaire de tout faire soi-même.

Dans un contexte où **la chaîne de valeur a de plus en plus tendance à se fragmenter**, savoir trouver les bons partenaires est capital. **Dans tous les cas, mieux on connaît sa donnée mieux on est à même de l'exploiter. Mieux on s'informe, mieux on est à même de retenir les bonnes offres.**

L'IA est amenée à transformer de nombreux secteurs d'activités, mais elle va également avoir besoin de nouveaux métiers. Elle ne pourra se développer qu'avec de nouvelles compétences humaines et de nouveaux savoir-faire. Le cloud public peut alors devenir un vecteur fort d'innovation à condition d'être correctement exploité par ses clients.

Je terminerai en remerciant une nouvelle fois tous les participants aux interviews ainsi que mon directeur de thèse professionnelle, monsieur Yoann Juge, « Responsable IT – Projets, Innovation & Développement » chez « OFI Asset Management ».

6 Annexes

6.1 INTERVIEWS

La forme et le rendu des interviews peut varier d'une personne à une autre fonction du mode d'échange retenu (face à face, téléphone, questionnaire rempli par la personne)

6.1.1 AWS : Julien Simon : Technical Evangelist Artificial Intelligence

Date : 05/07/2018

Identité de la personne l'interviewée

Nom : Simon

Prénom : Julien

Titre : AWS Technical Evangelist Artificial Intelligence

Type d'interview : par téléphone

- 1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*
 - Formation d'ingénieur en développement logiciel
 - 25 d'expérience dans l'informatique, d'abord dans l'industrie puis le web
 - Anciens postes marquants : Directeur technique de Pixmania et de Criteo
 - Travaille chez AWS depuis octobre 2015 en tant que « Technical Evangelist »
 - Travaille à promouvoir l'offre d'IA présente chez AWS, va à la rencontre des clients et aide à l'identification de leurs besoins en leur démontrant les capacités d'AWS à traiter les uses cases autour de l'IA

- 2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud*
 - Le cloud se caractérise avant tout par une infrastructure virtualisée mondiale et recouverte d'une couche d'API
 - C'est une offre de services identiques partout dans le monde permettant aux utilisateurs de déployer « programmatiquement » leur infrastructure informatique.

- 3) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA*
 - Construire des systèmes avec des comportements humains
 - Machine Learning : Systèmes capables d'apprendre des comportements sans être explicitement programmés grâce des données travaillées

(dataset) : suppose qu'un humain fournisse et traite le dataset pour en extraire les « features » (Data Scientist)

- DeepLearning : sous ensemble du machine Learning : extraction automatique des features (mais nécessite toujours un nettoyage préalable des données)
- Dataset : de l'ordre de plusieurs To (limite d'une base relationnelle)

Autres informations :

- Explosion du deep Learning à partir de 2006-2008
- 2012 : arrivée massive des GPU dans le cloud

4) *Êtes-vous fournisseur de solutions d'IA*

a. Si oui, quel est votre positionnement ? pouvez-vous nous faire une description de votre offre (marché, cible, missions, clients)

- Deux cibles
 - Clients déjà sachant sur l'IA et disposant de ses datasets et de ses algorithmes : AWS propose des solutions permettant la réutilisation de l'existant, mais en supprimant les contraintes liées à l'infrastructure
 - Clients cherchant à faire de l'IA : AWS propose des services clefs en main avec des API spécialisées et des modèles déjà prêts pour accélérer l'adoption de l'IA

b. Depuis combien de temps votre offre (ou vos offres) est-elle (sont-elles) disponible(s) ?

- Offre débutée en 2014 avec « Amazon Machine Learning » (ML sur étagère)
- Complétée début 2016 avec des API spécialisées dans le traitement de l'image et du texte
- Fin 2017 : arrivée de « SageMaker » plateforme entièrement gérée pour l'exécution d'algorithme d'IA

c. Votre offre va-t-elle évoluer ? Dans quelles directions ?

- Les offres AWS tournées vers l'IA vont continuer d'évoluer en s'améliorant itération après itération. AWS dispose de partenariat fort avec certains équipementiers spécialisés dans les processeurs et GPU pour disposer dès que possible de puces spécialisées et toujours plus performantes pour les algorithmes d'IA.

d. Quels freins rencontrez-vous aujourd'hui dans le développement de vos offres ?

- La mise en œuvre d'offres autour de l'IA nécessite beaucoup de compétence et de savoir-faire. AWS doit sans cesse recruter de nouveaux spécialistes sur le sujet qui ne sont pas toujours très nombreux.

e. Avez-vous des concurrents ? Si oui lesquels ?

- Les principaux autres fournisseurs de cloud public (Google, Azure, IBM...) proposent également des offres autour de l'IA

f. L'IA est-il un enjeu important pour vous ?

- C'est un marché en pleine expansion ou AWS souhaite proposer les offres les mieux adaptées et les plus performantes grâce à son infrastructure cloud connue et réputée.

5) Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ?

6) Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?

- Le cloud est particulièrement adapté à l'IA pour 3 raisons
 - Le stockage : l'IA a besoin de beaucoup d'espace pour stocker les datasets nécessaires aux algorithmes de Machine Learning et de DeepLearning
 - L'élasticité du compute : l'IA a des besoins forts en puissance processeurs, mais pas forcément tout le temps. De plus la mise à disposition de GPU optimise fortement l'efficacité des algorithmes d'IA
 - Les outils : AWS travaille avec de nombreux partenaires et éditeurs pour fournir des outils optimisés pour l'IA. On retrouve notamment des versions de bibliothèques OpenSource dont les performances sont grandement améliorées, car optimisées pour fonctionner sur l'infrastructure cloud du provider.

7) *Pensez-vous que l'IA va transformer des métiers ? Si oui lesquels et de quelles façons ?*

- L'IA va transformer certains métiers en allant jusqu'à remplacer certaines fonctions
- Certains process nécessiteront moins d'humains
- Les entreprises pourront traiter le cas de l'IA selon deux axes
 - Une vision « cost cut » où l'entreprise cherchera à tout prix à diminuer ses couts salariaux
 - Une vision à long terme où l'entreprise capitalisera sur ses valeurs pour faire évoluer son personnel vers d'autres taches plus centrées sur le business métier et l'innovation
- L'automatisation est une chose, mais ce n'est pas l'automatisation qui apporte l'innovation. L'IA fonctionne sur des projets muris.

8) *De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ?*

- La France est globalement en retard sur l'IA
- Nous avons une culture où l'on aime faire les choses soit même et où l'on préfère d'abord observer l'adoption d'une nouvelle technologie par d'autre avant de l'appliquer.

9) *Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?*

- La peur du changement : « il est urgent d'attendre »
- Le scepticisme
- L'explicabilité : c'est une vraie problématique, mais trop souvent mise en avant sur des cas de figure où elle ne s'applique pas. Elle est particulièrement rattachée au Deep Learning, mais le Deep Learning n'est pas le seul moyen de faire de l'IA

10) *Pensez-vous que l'IA a un intérêt/impact particulier pour les entreprises financières. Si oui pouvez-vous développer ?*

- De par leur activité, les entreprises financières ont toujours eu un besoin fort en analytique
- Ce sont des métiers où l'IA peut énormément apporter, notamment pour le calcul de risque et l'innovation.

11) Quelle va être l'évolution de l'IA et/ou du cloud pour vous ?

- Disparition des interfaces hommes-machines non vocales...
- Arrêter d'expliquer comment fonctionne un ordinateur (30 à 40% de gens sont mal à l'aise avec l'automatisation parce qu'ils n'arrivent pas à comprendre comment cela fonctionne, il va falloir imposer la méthode de communication naturelle de l'homme : la voix)
- mise à disposition dans le cloud de matériel spécialisé pour le traitement de l'IA (passage du GPU à de vraies puces spécifiquement conçues pour le deep learning (Intel Nervana))

6.1.2 CDO d'un groupe bancaire

Date : 27/08/2018

Identité de la personne l'interviewée

Nom : Anonyme

Prénom : Anonyme

Titre : Chief Digital Officer pour un groupe bancaire

Type d'interview : par téléphone

1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*

- CDO pour un groupe bancaire
- Travaille essentiellement en Europe
- En charge depuis plus de 2 ans de développer les nouveaux usages IT et le digital
- Rôle d'évangélisation autour des nouvelles technologies
- Gère des expérimentations autour des nouvelles technologies

2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud ?*

- Ensemble des services IaaS, PaaS et SaaS.
- Accès à des services à valeur ajoutée
- Facilite la collaboration et la construction de projet commun (avec des partenaires et/ou des clients)
- Bénéficier d'une standardisation des technologies

3) *Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel ?*

- 2008 : Bascule d'une application métier en solution SaaS
- 2009-2010 : Réflexion sur cloud interne
- Depuis 2010 : Déploiement d'un cloud interne pour standardiser l'hébergement des applications, accélérer les déploiements, proposer un IaaS interne et les workflows associés
- Depuis 2018 : ouverture vers des clouds publics comme AWS et Azure
- Objectif : 80% de l'architecture en cloud, qu'il s'agisse de cloud privé ou public. Le choix se fait au cas par cas fonction des contraintes réglementaires et des besoins.

4) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA*

- L'émulation du comportement humain

5) *Utilisez-vous ou prévoyez-vous d'utiliser une offre d'IA dans le cadre de votre activité professionnelle ?*

- Beaucoup d'expérimentations en cours avec quelques passages en production

a. *Si oui lesquelles et pourquoi ? Quels retours pouvez-vous en faire ? Cette offre repose-t-elle sur une solution cloud ?*

- RPA (Robotic Process Automation)
- Extraction de champs dans des formulaires
- Reconnaissance d'écriture manuscrite
- Appel à des services « SaaS » pour certains cas de KYC (Know your Customer) (obligation légale)
- Usage de ressources internes pour exploiter des algorithmes de machine learning
- Usage de brique open source

b. *L'IA est-il un enjeu pour vous ?*

- Enjeux portés par le digital, mais au même titre que d'autres sujets (blockchain, cloud...)

6) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ? Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?*

- On a souvent de fausses idées sur les besoins de puissance pour faire de l'IA. Certains « uses cases » ou algorithmes sont exploitables sans nécessiter l'usage d'un cloud.
- Le cloud permet néanmoins d'expérimenter l'IA plus simplement et surtout facilite les phases d'industrialisation souvent très lourdes.

7) *De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ?*

- Retard sur les Américains, mais beaucoup d'expérimentation en cours

8) *Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?*

Pour l'IA :

- Peur du manque de maîtrise
- Perte d'autonomie
- Problème réglementaire
- Retours pas toujours immédiats : besoin d'affiner les modèles
- Manque d'expertise humaine
- Disparitions des métiers à tâches répétitives (même impact que l'industrialisation)
- Coût des « POC » encore élevée malgré le cloud

Pour l'IA à travers le cloud (public) :

- Mêmes difficultés que pour le cloud
 - Difficulté de faire modifier les contrats des principaux providers
 - Difficulté d'obtenir des audits auprès des principaux providers
 - Contraintes réglementaires notamment sur la localisation des données
 - Besoin de travailler différemment
 - Évolution ou changement des besoins métiers
 - Problématique de la propriété des données :
La nature de certaines données n'a pas d'intérêt propre à son activité et peut être largement partagée et exploitée dans des solutions d'IA mutualisées (API speech to text par exemple)
Dans d'autres cas, vous possédez de la donnée qui vous est propre et vous seuls les avez. Dans ce cas, il faut être particulièrement prudent aux outils d'IA retenus (ex service de traduction spécifique pour des documents financiers)

9) *Pensez-vous que l'IA a un intérêt/impact particulier pour les entreprises financières. Si oui pouvez-vous développer ?*

- Les entreprises financières travaillent historiquement avec beaucoup de données.
 - ➔ La matière première de l'IA reste la donnée
- L'IA peut trouver de nombreux « use case » dans les banques

10) *Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.)*

- Multiplication dans la diversité des services cloud
- Poursuite du morçèlement de la chaîne de valeur avec l'apparition de plus en plus de Bpaas (Business Process As A service)
- Partenariat entre grands acteurs (consortium bancaire)

11) *Quelle va être l'évolution de l'IA pour vous ?*

- Nous ne sommes qu'au début
- Apparition de nouveaux cas d'usages

6.1.3 DataKlub : Tony Alibelli : CTO

Date : 03/09/2018

Identité de la personne l'interviewée

Nom : Alibelli

Prénom : Tony

Titre : CTO DataKlub

Type d'interview : par téléphone

- 1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*
 - CTO de Dataklub depuis 4 ans, une startup spécialisée dans la prestation de services et le conseil aux entreprises dans le bigdata, notamment via les solutions Splunk.
 - 15 ans d'expérience dans l'IT de finance de marché (Support, Gestion de projet, Sécurité...)
 - Travaille actuellement au développement d'une solution d' « Immersive Analytics » (interaction avec la donnée via la réalité augmentée)
 - Intervient également en tant qu'enseignant en université

- 2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud ?*
 - Une facilité d'accès à des ressources IT à la demande.

- 3) *Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel ?*
 - Exploitation d'offre IaaS chez AWS depuis 4 ans (location de VM et d'instance GPU)

- 4) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA ?*
 - Beaucoup d'amalgames « un peu tout et n'importe quoi »
 - Machine Learning : apprentissage sans réseau de neurones, basé sur des algorithmes types « Random Forest »
 - Deep Learning : usage de réseaux de neurones, très utilisé aujourd'hui pour la reconnaissance d'images

5) *Utilisez-vous ou prévoyez-vous d'utiliser une offre d'IA dans le cadre de votre activité professionnelle ?*

a. Si oui lesquelles et pourquoi ? Quels retours pouvez-vous en faire ? Cette offre repose-t-elle sur une solution cloud ?

Cas de « Splunk »

- Exploitation de machine learning au sein de la solution SIEM (Security Information and Event Management) de Splunk
- Analyse comportementale au sein de la solution de supervision associée à Splunk (ex : détection d'une consommation anormale de charge processeur)
- L'outil peut être installé chez le client ou dans le cloud

Développement interne

- Prototype de NLP (Natural Language Processing)
- Exploitation de GPU et ressources physiques pouvant être dans le cloud, mais pas d'usage des API dédiées ou des solutions « sur étagères » du cloud
- Exploitation du SDK « Tensor flow » et du langage NVIDIA « Cuda » pour l'exploitation des GPU (Membre du programme Nvidia Inception)

b. Si non pourquoi

- API des providers clouds parfois trop rigides : ex : début de développement avec « Alexa » d'Amazon puis bascule sur un modèle interne, car top de limitations rencontrées.
- API des providers pouvant changer de modèle de facturation (ex Google Maps)
- Construire son modèle d'IA uniquement avec les solutions « clef en main » des providers présente un risque de dépendances fortes

c. L'IA est-il un enjeu pour vous ?

- Enjeux importants sur la recherche et le développement. Besoins forts en reconnaissance d'images et dans le traitement du langage

6) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ?*

- Sans cloud pas d'IA
- L'IA nécessite des architectures correctement dimensionnées accessibles le plus souvent uniquement à de grosses structures, le cloud permet à tous de faire de l'IA.
- L'IA nécessite beaucoup de besoins collaboratifs. Pour faire évoluer les algorithmes, il faut permettre au plus grand nombre de pouvoir s'intéresser au sujet et y travailler. Le cloud facilite grandement la mise en œuvre de cette collaboration.

7) *Pensez-vous que l'IA va transformer des métiers ? Si oui lesquels et de quelles façons ?*

- Risque de fin des téléconseillers notamment dans les banques
- Risque de fin des humains aux caisses des magasins avec le paiement en ligne automatique (ex : ouverture des boutiques Amazon aux États unis)
- D'un autre côté, l'IA va créer de nouveaux besoins et amener à la création de nouveaux métiers

8) *De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ?*

- Beaucoup d'usages pour la lutte contre la fraude dans le milieu bancaire
- Usage dans l'IT Ops pour faire de la maintenance prédictive
- Usage dans le marketing pour le suivi client et le profilage
- Usage dans l'assistance à la personne
- Les Américains ont effectivement une avance sur le sujet (ex mise en œuvre de détection de fausses identités dans les aéroports), mais la France bénéficie d'un vivier fort en compétences, notamment grâce à notre enseignement universitaire sur les mathématiques. On note également une volonté politique de mettre en avant la « french tech »

9) *Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?*

- À l'identique des années 90 qui a connu un grand désaveu de l'IA : des résultats trop loin des attentes
- Trop de marketing : ressources vendues VS coûts réels.
- Risque d'amalgame et de mélange des genres
- Réseau de neurones : on peut expliquer le fonctionnement, mais pas le résonnement (rapidité de la machine sur l'être humain)
- Contraintes réglementaires (attention, la loi ne doit pas freiner l'innovation)

10) *Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.)*

- Développement de plus en plus réalisé sur des briques propriétaires des clouds providers (architecture micro service)
- Besoins de plus en plus forts de communication intercloud

11) Quelle va être l'évolution de l'IA pour vous ?

- Nous ne sommes qu'au début
- Le rythme d'avancé risque d'être irrégulier
- Arrivée des ordinateurs quantiques : révolution dans les capacités de calcul
- « Il vaut mieux être acteurs que spectateurs »
- Pourquoi ne pas imaginer une IA capable de déplacer une architecture IT d'un cloud à un autre fonction des opportunités et des coûts

6.1.4 Google : Jean-Luc Moisan : Responsable Avant-Vente Services Financier

Date : 09/08/2018

Identité de la personne l'interviewée

Nom : Moisan

Prénom : Jean-luc

Titre : Responsable Avant-Vente – Services Financier Google

Type d'interview : par téléphone

1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*

- Customer Engineer chez Google – Secteur Finance France depuis décembre 2017
- Responsable d'une équipe de 7 personnes
- Ancien « Principal Sales Consultant » chez Oracle
- Rôles
 - Connaître les produits proposés par Google Cloud
 - Connaître les produits concurrents et rester informé des nouveautés
 - Connaître les besoins du client pour l'accompagner dans son projet

2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud ?*

- Un service cloud s'inscrit dans au moins l'une des trois catégories suivantes : le IaaS, le SaaS ou le PaaS (et bientôt d'autres)
- Un service cloud se caractérise par deux spécificités
 - Financière : offre disponible à la demande, sans engagement
 - Technique : offre une « configurabilité » importante pour correspondre à de nombreux cas d'usage

3) *Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel.*

- Travaille avec et sur le cloud depuis plusieurs années
- Création d'une société en 2009 reposant sur un modèle SaaS

4) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA ?*

- Tous les domaines logiciels qui couvrent les besoins finaux ou de développement pour fournir des outils de prédiction mathématique
- Machine Learning : ensemble de techniques qui rend le logiciel autonome dans ses calculs, peut être simple ou complexe, mais fournit toujours une réponse « probabilistique »
- Le DeepLearning est une évolution du machine Learning ou l'apprentissage passe par une topologie de réseau de neurones
- Reinforcement Learning : on ne supervise plus l'algorithme en écrivant en amont la prédiction attendue.

5) *Êtes-vous fournisseur de solutions d'IA*

- a. Si oui, quel est votre positionnement ? pouvez-vous nous faire une description de votre offre (marché, cible, missions, clients)*
- b. Depuis combien de temps votre offre (ou vos offres) est-elle (sont-elles) disponible(s) ?*
- c. Si oui, vos solutions reposent elles sur une architecture cloud si oui pourquoi ? Si non pourquoi ?*
- d. Votre offre va-t-elle évoluer ? Dans quelles directions ?*
- e. Quels freins rencontrez-vous aujourd'hui dans le développement de vos offres ?*
- f. Avez-vous des concurrents ? Si oui lesquels ?*
- g. L'IA est-il un enjeu important pour vous ?*

- L'IA est industrialisée depuis plusieurs années chez Google
- Plus aucun service grand public ne tourne sans IA (Gmail, Google Photos, Maps...)
- 4000 modèles sont actuellement en production chez Google : l'IA est un enjeu stratégique pour l'entreprise

- À travers son offre cloud, Google propose trois approches possibles pour ses clients :
 - Approche pour clients avancés sur l'IA (cas des grands réseaux bancaires). Ce sont des entreprises qui ont les talents nécessaires pour créer leurs propres modèles. Pour eux, Google met à disposition de l'infrastructure comme des CPU, des GPU, mais aussi des TPU spécialisés pour la mise en place de réseau de neurones. Google propose aussi une librairie OpenSource, TensorFlow, offrant un outil de programmation dédié à l'IA. Google propose également un réseau de stockage avancé offrant de très grosses capacités. (On ne conçoit pas un modèle de traitement d'image avec 50 images sources, mais avec des milliers de clichés)
 - Approche pour les clients ayant des besoins forts d'IA, mais pas de talent en interne et souffrant d'un manque d'expertise sur le sujet. Google fournit depuis peu un nouveau service baptisé « AutoML »(unique sur le marché) qui aide à la construction de son propre modèle. Le client n'a plus qu'à travailler et fournir son jeu de données. Pour l'instant disponible uniquement sur de l'analyse d'image, mais prochainement décliné sur d'autres domaines
 - Approche pour les clients ayant peu de compétence et de temps à consacrer à l'IA, mais avec des besoins simples et génériques. Google fournit l'accès à des API spécialisées (text to speech, traduction, traitement d'image...) pour obtenir des solutions « clef en main »

Dans les faits on s'aperçoit que les clients mixent les approches.

- L'IA est un marché ultra concurrentiel, mais Google se démarque par sa maîtrise de l'industrialisation de l'IA et son expérience acquise sur le sujet à travers ses outils et besoins internes.
- Google investit massivement dans la recherche (achat de DeepMind), car si le passage d'un modèle théorique à un modèle pratique en production est bien maîtrisé, il y'a maintenant besoin de faire évoluer les algorithmes et trouver de nouveaux principes mathématiques.

6) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ?*

- L'IA a été un sujet en « veille » pendant un temps par manque de puissance et le besoin de lancer des calculs sur des masses de données. Le cloud a permis de débloquent cette situation en proposant l'accès à la demande et sans besoin d'achat à long terme d'infrastructures élastiques et ultras performantes.

7) *Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?*

- Aucun responsable d'entreprise ne peut se permettre d'allouer de l'argent pour tester une IA type « Deep Learning » sans garantie de résultats. Le modèle économique du cloud permet de tester et de s'essayer à l'IA sans imposer d'investissement lourd de la part des entreprises.
- 8) *De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ? Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?*
- Globalement il y'a énormément de chantiers en cours sur l'IA dans les entreprises, mais certains par effets de modes avec peu de risque de passage en production
 - En France, il y'a nécessité de réinvestir dans des talents pour exploiter correctement l'IA et créer de la valeur.
 - Pas d'acteur français, d'où l'obligation de se tourner vers des entreprises principalement américaines et bientôt chinoises
 - L'explicabilité peut être un problème, mais peu souvent se contourner :
 - ex Détection de fraude par HSBC : la loi impose de pouvoir expliquer le rejet d'une transaction, l'IA est donc exploitée comme un outil d'aide à la décision. HSBC a laissé son service de détection de fraude en place et ne l'a pas remplacé par une IA, cette dernière permet d'accélérer les prises de décision, le choix final restant à l'humain qui peut donc expliquer son raisonnement.

Une photo dans « Google Photos » passe par 64 filtres avec une analyse par itération par réseau de neurones imbriqués. Dans ce cas, il est très compliqué d'expliquer le cheminement d'analyse de l'image, mais ce n'est pas un problème dans ce cas d'usage.

- 9) *Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.) Quelle va être l'évolution de l'IA pour vous ?*
- Avoir de l'IA spécialisée et très orientée métier :
Ex : la compréhension du mot « Livret A » n'a pas le même sens dans le domaine financier, littéraire ou médical. C'est la limite des IA entraînées sur des modèles génériques aujourd'hui.
 - L'IA va avoir de plus en plus d'impact pour couvrir les besoins de conformité et d'audibilité. C'est un sujet important pour les banques (éviter les affaires "Jérôme Kerviel").

6.1.5 Nephelai : Romain Mangeret : Cofondateur

Date : 28/06/2018

Identité de la personne l'interviewée

Nom : Mangeret

Prénom : Romain

Titre : Cofondateur de Nephelai

Type d'interview : document rempli par la personne

1) Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?

Je suis Romain Mangeret, cofondateur de Nephelai, fintech qui développe une solution d'assistance à la saisie de transactions financières s'appuyant sur l'IA.

Je suis ingénieur de l'École Polytechnique. J'ai travaillé 10 ans chez l'éditeur de logiciel financier Sophis, ensuite devenu Misys puis Finastra au fil des acquisitions et fusions. J'étais initialement consultant Front Office, en charge de l'implémentation du logiciel Sophis Value chez les sociétés de gestion. J'ai ensuite occupé le poste de consultant avant-vente. Finalement je suis devenu chef de produit, tout d'abord des outils Front Office pour les sociétés de gestion, puis plus généralement de la solution de Misys pour les sociétés de gestion, appelée FusionInvest.

Il y a 8 mois, j'ai quitté Misys/Finastra pour fonder Nephelai avec 4 associés.

Nephelai est à la fois consommateur de cloud et fournisseur de solution d'IA.

2) Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud ?

Le cloud est pour moi un outil qui offre des possibilités techniques peu accessibles pour des petites structures ou des particuliers. Plus précisément, ces possibilités sont l'accès à plusieurs machines qui permettent de paralléliser les calculs, la réplication automatique des données qui évite les pertes en cas de problème technique, l'élasticité c'est-à-dire la capacité à réquisitionner un grand nombre de machines de façon ponctuelle en cas de pic de charge.

3) Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel ?

Nous utilisons le cloud depuis le début de Nephelai il y a 8 mois. Nous utilisons évidemment plusieurs solutions qui s'appuient elles-mêmes sur le cloud (gmail, JIRA, etc.), et utilisons directement différents clouds pour faire fonctionner la solution que nous vendons.

b. Si vous l'exploitez, quels services utilisez-vous ? Pourquoi ?

Notre solution s'appuie sur des algorithmes d'IA qui sont très consommateurs en ressources de calcul. Le cloud offre une solution tout à fait adaptée à ce problème.

Le cloud pose néanmoins certains problèmes dans son utilisation.

La sécurité des données est une préoccupation capitale pour nos clients, et certains refusent qu'elles soient envoyées sur un cloud.

Son utilisation peut être très coûteuse. Ceci dépend fortement des caractéristiques techniques de chaque machine (une unique machine très puissante coûte beaucoup plus cher que plusieurs machines communes), et du niveau de priorité pour l'accès à la machine. Pour optimiser le coût d'utilisation, il convient de concevoir des algorithmes fortement parallélisables, ce qui permet d'utiliser plusieurs machines très peu coûteuses, et en outre de limiter chaque calcul à un process court, donc l'usage de la machine même en priorité faible risque peu d'être préempté.

4) Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA.

Dans son usage courant, l'IA est un terme générique qui recouvre 2 techniques assez différentes :

- L'apprentissage statistique (Machine Learning)
- L'apprentissage profond (deep learning), qui s'appuie sur des réseaux de neurones avec plusieurs couches cachées

L'apprentissage statistique comprend des méthodes très variées, de simples régressions linéaires à des arbres de décision complexes de type Random Forest. Ces méthodes s'appliquent bien sur des données structurées (par exemple les transactions financières). L'apprentissage profond nécessite en général des ressources beaucoup plus importantes pour être mis en œuvre. Contrairement au ML, il s'applique bien aux données non structurées (par exemple la reconnaissance d'image).

La particularité de l'IA est que la solution est spécifique à chaque problème, et le succès dépend fortement de la capacité de l'équipe à intégrer les connaissances métiers dans la mise en œuvre des algorithmes.

6) Êtes-vous fournisseur de solutions d'IA

Oui, l'IA est au cœur de notre offre. Notre solution a vocation à détecter les erreurs de saisie sur les transactions financières. L'IA donne d'excellents résultats pour 3 raisons :

- Les données sont naturellement structurées (les transactions sont des formulaires)
- L'accès à un historique profond est simple, car toutes les transactions sont historisées et les modifications auditées (ce qui permet l'utilisation de méthodes supervisées)
- Il existe des schémas statistiques sous-jacents à ces transactions (structure du marché, habitude du trader, etc.).

Comme expliqué précédemment, notre offre s'appuie sur le cloud.

- a. Quel est votre positionnement ? pouvez-vous nous faire une description de votre offre (marché, cible, missions, clients)*

Nous offrons une solution d'assistance à la saisie d'ordres et de transactions financières, et de détection d'erreur sur ces mêmes transactions. Notre solution s'adresse à toute institution financière traitant sur les marchés de capitaux avec des volumes significatifs (banque d'investissement, société de gestion, industriel, asset servicers, etc.).

Notre but est d'aider nos clients à améliorer leur efficacité opérationnelle, et réduire le coût de transaction.

- a. Depuis combien de temps votre offre (ou vos offres) est-elle (sont-elles) disponible(s) ?*

Notre société existe depuis 4 mois.

- b. Si oui, vos solutions reposent-elles sur une architecture cloud si oui pourquoi ? Si non pourquoi ?*

Notre solution s'appuie sur le cloud (voir les explications précédentes).

- c. Votre offre va-t-elle évoluer ? Dans quelles directions ?*

Très certainement, mais il est un peu tôt pour connaître la direction.

- d. Quels freins rencontrez-vous aujourd'hui dans le développement de vos offres ?
Technique : outils coûteux, complexes ou non disponibles pour votre activité ?*

Le coût peut être élevé certes. Ce n'est pas un frein, car il existe des moyens de le réduire fortement. Comme je l'avais expliqué ailleurs dans le doc, si l'on se contente de machines unitaires pas très puissantes et que l'on accepte de risquer de se faire préempter la machine, le prix est réduit d'environ 80% chez certains fournisseurs de cloud (en fait on utilise simplement les machines non utilisées à un instant donné). Par contre, pour que ça fonctionne il faut que les algo puissent être parallélisés sur de nombreuses machines peu puissantes, et découpés en job aussi courts que possibles pour limiter le risque de voir la machine préemptée.

L'explicabilité est un problème fondamental pour nous. Pas pour des raisons juridiques (en tout cas nous n'avons jamais eu ce genre d'objection pour l'instant), mais tout simplement pour exploiter les résultats. Notre objectif est de prédire si une transaction est erronée à la saisie ou non. Or une transaction peut comporter plus d'une centaine de champs. Si nous nous contentons d'un résultat sur la transaction, ça laisse beaucoup de travail à l'opérateur pour trouver l'erreur. Il est nécessaire d'avoir des détails sur la raison de l'alerte, comme le champ ou la combinaison de champs responsables de la levée de l'alerte. Pour cette raison nous avons dû développer nos propres algorithmes, car aucun algorithme ouvert n'est performant sur le sujet et ne permet d'obtenir des explications.

C'est une grosse faiblesse de beaucoup d'algo d'IA, et du deep learning en particulier, qui fonctionnent comme des boites noires.

e. Avez-vous des concurrents ? Si oui lesquels ?

Nous n'avons pour l'instant pas de concurrents directs, nous sommes les seuls à travailler précisément sur les erreurs de saisie dans les transactions financières (idée originale de notre fondatrice et présidente).

Beaucoup d'entreprises travaillent sur des solutions génériques d'IA (par exemple IBM Watson) qui auraient vocation à être appliquées facilement sur de nombreux problèmes, et sont une forme de concurrence.

f. L'IA est-il un enjeu important pour vous ?

Oui, crucial.

7) Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ?

Le cloud est outil indispensable pour de petites structures comme la nôtre qui souhaitent mettre en œuvre des solutions d'IA qui nécessitent d'importants moyens de calculs.

8) Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?

Indubitablement.

9) Pensez-vous que l'IA va transformer des métiers ? Si oui lesquels et de quelles façons ?

L'IA va très certainement transformer en profondeur de nombreux métiers, on trouve suffisamment de points de vue sur le sujet dans la presse !

10) De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ?

Je pense que l'application de l'IA est encore marginale par rapport à son potentiel dans la majorité des métiers.

11) Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?

Je pense que l'un des plus gros obstacles est le manque de compréhension de l'IA.

D'un côté, elle est souvent présentée comme un outil magique qui pourrait régler tous les problèmes, ce qui inévitablement aboutit à une forte déception lors de la mise en œuvre.

D'un autre côté, certaines personnes peuvent se sentir menacées, d'un point de vue professionnel, par l'IA et en conséquence s'opposer à son déploiement.

12) Pensez-vous que l'IA a un intérêt/impact particulier pour les entreprises financières. Si oui pouvez-vous développer ?

L'industrie financière est un terrain idéal d'expérimentation de l'IA, puisqu'elle offre un accès aisé à de très profonds historiques de données structurées. En outre, il existe une vraie culture scientifique et d'innovation dans la finance de marché (la vague précédente était celle du calcul stochastique et de ses quants) qui rend d'autant plus propice son développement rapide.

Il est évident que l'IA aura des impacts profonds sur de nombreux métiers de la finance, c'est déjà le cas depuis plusieurs années avec le trading automatique par exemple.

La connaissance des clients est probablement un des domaines les plus prometteurs pour l'IA, conjointement avec l'avènement de plateformes de distribution blockchain, qui suppriment les intermédiaires et permettent un accès plus direct à l'information sur les clients.

13) Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.)

14) Quelle va être l'évolution de l'IA pour vous ?

Time will tell !

6.1.6 Orange : Julien Bermudez : Business Manager Cloud For Business

Date : 29/08/2018

Identité de la personne l'interviewée

Nom : Bermudez

Prénom : Julien

Titre : Business Manager "Orange Cloud for Business"

Type d'interview: face à face

1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*

- Business Manager "Orange Cloud for Business" depuis 2016
- 2012-2016 : Architect IT Cloud chez Orange
- Chez Orange depuis 2009, a débuté sur le réseau FTH WAN, puis sur le support du service de messagerie managé via une offre Exchange.
- Rôle : Promouvoir le cloud Orange, proposer des services adaptés aux demandes clients et les accompagner dans leur migration.

2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud ?*

- Le cloud est un modèle économique avant tout.
- Très grossièrement, techniquement, c'est la mise à disposition d'hyperviseur en libre-service.
- C'est le passage d'un modèle CAPEX vers un modèle OPEX pour les entreprises pour leurs usages IT

3) *Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel ?*

- L'IA regroupe énormément de choses comme les réseaux d'apprentissages, les « chat bots » et d'autres technologies. Tout dépend du contexte et des tâches exécutées.

4) Êtes-vous fournisseur de solutions d'IA ou de cloud ?

- Orange dispose de 7 datacentres en France pour l'hébergement de son cloud
- Orange propose des services IaaS reposant soit sur une plateforme VMware ou une plateforme OpenStack (partenariat avec Huawei)
- Orange ne propose pas d'offre spécifique IA, mais met à disposition une infrastructure IaaS optimisée pour l'IA, avec par exemple, la location d'instance GPU
- Une offre est en cours d'élaboration appelée MLS pour « Machine Learning Service » de manière à simplifier l'accès au machine Learning aux entreprises qui le souhaitent.
- Orange propose des accompagnements sur-mesure, de bout en bout, pour construire sur son infrastructure IaaS des services SaaS pour ses clients. Orange a par exemple mis en œuvre un chat bot pour la « Caisse Nationale Vieillesse » permettant d'améliorer le parcours utilisateur dans la préparation de sa retraite.

L'IA est-il un enjeu pour vous ?

- La priorité est donnée à l'agrandissement de l'offre Cloud. L'IA reste pour l'instant des offres en devenir.

5) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ? Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?*

- L'IA a déjà été déployée sans cloud (ex cas d'IBM)
- Le cloud reste avant tout un avantage économique pour le déploiement et l'expérimentation de solutions d'IA.

6) *Pensez-vous que l'IA va transformer des métiers ? Si oui lesquels et de quelles façons ?*

- Certains métiers vont profondément être impactés par l'IA, par exemple les métiers liés au « télé conseil », mais également dans le transport (technologies de plus en plus nombreuses dans les camions par exemple)
- L'IA va avoir le même impact que la révolution industrielle pour les métiers à faibles valeurs ajoutées.

7) *De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ?*

- On note un retard par rapport aux Etats Unis, mais nous ne sommes qu'au début de l'IA.
- On note une très forte concurrence entre les grands providers de clouds. Il y'a donc un enjeu à grossir vite et à forger des alliances.

8) *Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?*

- Difficulté à trouver des clients avec des uses cases pertinents
- Peur de l'usage de l'IA

9) *Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.)*

- Beaucoup de consolidation (rachat de petites structures par des plus grosses)
- Des offres d'infogérance incluses dans les services Cloud

10) *Quelle va être l'évolution de l'IA pour vous ?*

- Développement de plus en plus de services.

6.1.7 QuantMetry : Aurelia Nègre : Data Scientist

Date : 01/08/2018

Identité de la personne l'interviewée

Nom : Aurélie

Prénom : Nègre

Titre : Data Scientist

Type d'interview : Face à face

1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*

- Data Scientist chez Quantmetry depuis 2 ans
- Quantitative Analyst pendant 2 ans à la Banque de France.
- Rôle actuel : « Faire parler la data »
 - Accompagne les clients pour l'analyse et à la gestion de leurs données.
 - Conçoit et aide au déploiement des algorithmes permettant le traitement des données fonction d'un but recherché (identifier des erreurs de saisies, alerter sur des anomalies, remonter une information critique, alimenter un tableau de bord...)
 - Développer des briques d'analyses
- Très intéressée par l'univers financier et la défense

Remarques marquantes durant l'interview

- Un data analyste est souvent plus orienté métier qu'un data scientist. Un data scientist peut s'appuyer sur des algorithmes existants capables de s'adapter à différentes situations et contextes (modulo des adaptations et un partenariat avec les métiers)
- Un data scientist est en mesure d'indiquer quels outils et moyens IT sont nécessaires à l'exécution des modèles algorithmiques qu'il conçoit.
- On observe une fusion des cultures entre les métiers en charge de la data dans l'IT et les statisticiens : réalisent parfois des opérations de mêmes types, mais exploitent un vocabulaire différent

2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud*

- Accès à des ressources de stockage et de calcul de façon simple et rapide
- Elasticité des infrastructures proposées

3) *Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel.*

- Peu de cas d'usage pour l'instant
- Usage de VMs et de stockage principalement chez AWS

4) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ?*

- IA : difficile de donner une définition unique, mais pourrait grossièrement se résumer à tout système qui visent à reproduire un raisonnement.
- Beaucoup de « buzz » sur l'IA, mais peu de réelles nouveautés sur le sujet, les grands principes souvent associés existent depuis longtemps, mais n'avaient pas trouvé de cas d'usages (ou fonctionnaient mal) par manque de données et de capacité de traitement
 - IA symbolique : type « If/Else » ou arbre de décision
 - IA connexionniste : Concept de « Réseau de neurones »
DeepLearning : principe datant des années 80
 - Machine Learning : type « Random Forest » : existe depuis 2001, analyse de donnée empirique

5) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ? Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?*

- Pas de lien direct : comme cité précédemment les concepts et les algorithmes derrière l'IA ont été conçus et pensés avant l'avènement du cloud
- Au même titre qu'un simple ordinateur, le cloud est avant tout une ressource pour l'exploitation de l'IA, mais offre toute son élasticité et sa capacité de stockage pour traiter les afflux massifs d'informations qui rendent aujourd'hui l'IA pertinente.

Remarques marquantes durant l'interview

- Contrainte du cloud : faire sortir sa donnée. Peu potentiellement être un frein pour certaines entreprises ou certains domaines d'activités, dont la finance.
- Méfiance des solutions clouds d'IA « clef en main » : difficultés de mesurer le risque des algorithmes employés et surtout de la fiabilité des données de sortie
 - Ex : Reconnaissance de race canine d'après une photo : les images de « Husky » présentaient l'animal dans un paysage neigeux : la neige est devenue pour la machine un critère de sélection pour identifier un chien comme un « Husky »
- Coût de stockage devenu particulièrement bas : répond à la contrainte de l'explosion de la donnée

6) *Pensez-vous que l'IA va transformer des métiers ? Si oui lesquels et de quelles façons ?*

- Des emplois vont certainement disparaître notamment dans l'univers administratif, mais d'autres besoins apparaîtront pour donner lieu à de nouveaux métiers.

7) *Quels pourraient être les freins à l'adoption de l'IA dans le monde professionnel ?*

- L'explicabilité : comprendre pourquoi ça marche (ou pas) ? : tâche pouvant se révéler difficile, mais sur laquelle beaucoup de gens travaillent
 - Approche « Lime » : Approximer un modèle complexe non interprétable par un modèle interprétable (par exemple un modèle linéaire)
- En complément : annexe 71 RGPD « La personne concernée devrait avoir le droit de ne pas faire l'objet d'une décision, qui peut comprendre une mesure, impliquant l'évaluation de certains aspects personnels la concernant, qui est prise sur le seul fondement d'un traitement automatisé et qui produit des effets juridiques la concernant ou qui, de façon similaire, l'affecte de manière significative, tel que le rejet automatique d'une demande de crédit en ligne ou des pratiques de recrutement en ligne sans aucune intervention humaine »

Remarques marquantes durant l'interview

Les banques ont besoin d'évaluer le risque d'un modèle pour pouvoir le retenir en production. C'est une démarche encore complexe aujourd'hui.

8) *Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.) Quelle va être l'évolution de l'IA pour vous ?*

- IA : Nous ne sommes qu'au début de l'histoire
- Cloud : va continuer à se développer et devenir presque un standard pour toute entreprise
- Marché fleurissant des GPU
- Arrivée de Loihi : premier processeur neuromorphique spécialement conçu pour reproduire un réseau de neurones.

Remarques marquantes durant l'interview

- Le métier de Data Scientist risque fort d'évoluer : aujourd'hui travaille essentiellement sur la notion de « Build » demain sans doute plus orienté vers la notion de « Run »

6.1.8 **Quinten** : Anna Nesvijevskaia Directeur Data Strategy et Thomas Schott
Directeur technique

Date : 07/08/2018

Identité de la personne l'interviewée

Nom : Nesvijevskaia
Prénom : Anna
Titre : Directeur Data Strategy chez Quinten

Nom : Schott
Prénom : Thomas
Titre : Directeur technique chez Quinten

Type d'interview : Face à face

- 1) *Pouvez-vous décrire votre entreprise et votre offre*
 - Quinten est une société de conseil qui accompagne ses clients sur la génération de valeur par sa donnée.
 - Quinten dispose d'une expertise métier dans le domaine de la médecine, de la banque, de l'assurance et du web
 - C'est le besoin client qui prime avant tout : importance du but recherché
 - Identification du potentiel des données du client

- 2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud ?*
 - Un ensemble de ressource élastique permettant d'aller sur de nouveaux usages et saisir de nouvelles opportunités.

- 3) *Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel ?*
 - Peu d'usage du cloud pour la production actuelle, mais utilisations de technologies communes, notamment sur les outils de déploiement et de calculs
 - Forte veille technologique sur le cloud
 - Essais en cours ou à venir de certaines offres cloud (notamment avec AWS)

4) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA ?*

- IA est un terme très vague, le plus souvent utilisé à des fins de communication/marketing qu'on peut potentiellement rattacher à n'importe quels types d'algorithmes.
Algorithme = considère 1 ou plusieurs « input » + réalise un traitement fini non ambigu + donne un résultat
- D'un point de vue scientifique/technique, on préfère parler du nom/type d'un algorithme plutôt que d'IA.
- Ce qui vient caractériser un algorithme d'IA c'est surtout son « input ». On vient injecter un historique de données dans le but de modéliser une fonction d'apprentissage d'un être humain.
- « IA » est devenue un « phénomène ».
- Le Deep Learning est un sous ensemble du Machine Learning

5) *Utilisez-vous ou prévoyez-vous d'utiliser une offre d'IA dans le cadre de votre activité professionnelle ?*

- Potentiellement oui, mais cela va dépendre des usages clients et des besoins.
- Pas d'usage d'API d'IA « clef en main » pour l'instant, car Quinten dispose de compétences en interne pour ses besoins. Pas de bridage rencontré sur les algorithmes internes, mais des réflexions sont en cours sur l'usage de solutions cloud.

6) *Êtes-vous fournisseur de solutions d'IA. Quelle approche avez-vous ?*

- Sur les besoins IA, l'approche conseillée est plutôt d'établir d'abord une infrastructure bigdata pour ensuite y connecter des solutions d'IA
- Quinten conçoit et propose son propre algorithme de traitement de la donnée
→ Permet de trouver le bon niveau d'abstraction

7) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ?*

- L'IA a beaucoup été mise en avant par les principaux providers de cloud
- Comme pour le bigdata, les acteurs du cloud ont alimenté « le buzz »
- Beaucoup d'entreprises traitaient déjà beaucoup de données avec des solutions internes (cas des assurances par exemple)
- Avec l'IA, le cloud s'inscrit dans la continuité de ce qu'il a pu proposer et continue de proposer sur le bigdata : permettre d'arrêter de perdre de la donnée, ce qui s'avère pratique, mais attention aux investissements à l'aveugle.

8) *Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?*

- Grâce à son élasticité et ses performances, le cloud est naturellement adapté à traiter des problématiques d'IA.
- Le cloud permet l'usage d'algorithmes connus (parfois depuis très longtemps) qui pouvaient jusqu'à présent se montrer complexes à mettre en œuvre de par la quantité de données et de puissance nécessaires pour les rendre efficaces. On ouvre ainsi l'éventail de choix des algorithmes.
(Il est néanmoins à noter que des algorithmes simples existent, sans forcément nécessiter de grosses puissances de calcul)
- Il faut distinguer la phase d'exploration de la phase d'industrialisation.
- Associer l'IA et le cloud est tout fait possible, mais cela dépend avant tout des entreprises.
 - ➔ Des sociétés qui traitent leurs données de façon historique peuvent aller vers l'IA par intérêt pour les nouveaux modes de traitements de la donnée. Ils peuvent alors se tourner vers le cloud pour s'essayer à ces nouveaux usages.
- Le cloud, grâce au service managé d'IA qu'il peut proposer permet de s'adresser à plus de monde et banaliser l'usage de certains algorithmes, mais avec le risque de mal les utiliser pour les clients.

9) *Pensez-vous que l'IA va transformer des métiers ? Si oui lesquels et de quelles façons ?*

- Va libérer plus d'espace pour l'expertise et la prise de décision
- Va permettre de responsabiliser davantage les personnes (l'IA va fournir des outils complémentaires qui vont limiter les excuses possibles d'une personne pour ne pas pouvoir réaliser ses objectifs)
- L'IA va avoir les mêmes effets que l'industrialisation pour prendre en charge les tâches répétitives
- Mise en œuvre d'outils d'aide à l'investigation : inclus la notion d'accélération et donc de transformations des métiers, notamment dans l'analytique.

10) Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?

- Retour sur investissement parfois difficile
 - ➔ Les provider de cloud peuvent parfois générer beaucoup d'idées reçues qui se montrent fausses si l'entreprise qui souhaite faire de l'IA n'est pas suffisamment mature sur ses besoins
- Échec de mise en application si pas d'autres motivations que la pression des éditeurs de clouds.
- Explicabilité des résultats : peut être un frein en fonction des métiers et des usages concernés. Parfois on cherche juste de la précision.
- Il y'a encore une maturité instable autour de l'IA : attentes utopiques VS méfiance au plus haut point. (D'où l'importance de l'accompagnement au changement)
- Volonté de garder ses données en interne (mais les critères évoluent)
- Avoir beaucoup de ressources en amont peut impliquer un certain laxisme dans les phases de mise en œuvre traduisant des résultats décevants
 - ➔ On peut être tenté de faire n'importe quoi

6.1.9 SESAmm : Sylvain Forte : Cofondateur et CEO

Date : 03/08/2018

Identité de la personne l'interviewée

Nom : Forte

Prénom : Sylvain

Titre : Cofondateur et CEO de SESAmm

Type d'interview : téléphonique

1) Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?

- Cofondateur et CEO de SESAmm
- Ancien Ingénieur informatique

2) Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud

- Principalement une réserve de puissance de calcul

3) Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel.

- Utilisateur confirmé du cloud
- Sesamm exploite des ressources AWS
- Beaucoup de veilles sur les technologies cloud et sur les offres disponibles

b. Si vous l'exploitez, quels services utilisez-vous ? Pourquoi ?

- Ressource IaaS principalement
- Usage de CPU et de GPU pour du calcul Spark et de l'entraînement de modèle
- Usage de framework tensorflow
- Pas d'exploitation d'outils type « API » spécialisées : la plupart des outils d'AI « clef en main » reposent sur des modèles qui se sont construits sur des données génériques, souvent éloignées des données propres à la finance.

4) Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ?

- Globalement l'IA essaye de simuler une intelligence humaine
- L'IA englobe beaucoup de technologies et d'algorithmes
 - Ex : NLP : Natural Language Processing : héritage de la linguistique informatique pour modéliser le langage humain
 - Le Machine Learning dans lequel on retrouve un sous-domaine qui est le Deep Learning
 - Machine Learning : mise à disposition de l'algorithme des variables de sélections

- Deep Learning : modèles plus complexes que les modèles de machines Learning. L'algorithme identifie lui-même les caractères différentiant.
- Le « Reinforcement Learning » utilisé par exemple dans les systèmes de jeux.

6) Êtes-vous fournisseur de solutions d'IA

- Solution de traitement de l'information exploitant des algorithmes de NPL pour fournir des indices comportementaux à destination des Asset Manager
 - 250 000 sources textuelles (Forum, Trading social, réseaux sociaux, communauté reddit...)
- b. Quel est votre positionnement ? pouvez-vous nous faire une description de votre offre (marché, cible, missions, clients)**
- Cible principalement les sociétés d'Asset Management (partenariat en cours avec La Française, Candriam)
 - Fournir des solutions permettant l'établissement de stratégies différenciantes pour les Asset Manager
 - Deux offres disponibles
 - Accès aux données traitées via interface web et API
 - Offres projets pour la création de modules de prise de décisions.
- g. Si oui, vos solutions reposent-elles sur une architecture cloud si oui pourquoi ? Si non pourquoi ?**
- Les architectures cloud sont exploitées pour l'entraînement des modèles, car c'est une opération qui nécessite beaucoup de ressources mais de façon ponctuelle. Une fois les modèles disponibles, ceux-ci ont plutôt tendance à tourner hors cloud public pour des questions de sécurité et de souhaits client. (Machine Bare Metal)
- h. Votre offre va-t-elle évoluer ? Dans quelles directions ?**
- Devenir « l'IBM de la finance »

i. *Quels freins rencontrez-vous aujourd'hui dans le développement de vos offres ?*

Technique : outils couteux, complexes ou non disponibles pour votre activité ?

- L'engouement est très fort pour la solution Sesamm. Le marché financier est très demandeur d'IA
- Il peut y'avoir de l'incompréhension de certains clients sur les technologies liées à l'IA et sur leurs capacités à répondre à certains uses cases.
- L'explicabilité n'est pas vraiment un problème dans le domaine de l'Asset Management et de l'usage d'algorithmes de NPL. Les logiques utilisées peuvent être expliquées. Il ne s'agit pas de fournir des modèles d'investissement ou le besoin d'explicabilité est beaucoup plus fort.

j. *L'IA est-il un enjeu important pour vous ?*

- Cœur de l'offre Sesamm

7) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ?*

- Sans cloud, pas de puissance et pas de puissance, peu d'IA
- 90 % des startups exploitant de l'IA sont dans le cloud

10) *De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ?*

- Beaucoup de retard en France
- L'Europe dispose de datalab avec des data scientist opérationnels, mais les technologies d'IA sont encore très mal connues
- Le marché américain est beaucoup plus mature et dispose déjà d'outils pour tester et challenger les offres d'IA

13) *Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.)*

- Cloud décentralisé : usage de machines réparties partout dans le monde (ex : blockchain), les principaux providers du cloud devant faire face à de plus en plus de problématiques énergétiques.

14) *Quelle va être l'évolution de l'IA pour vous ?*

- « One Shot Learning » : besoin de moins de données et de moins d'itérations pour construire un modèle et se rapprocher ainsi plus du fonctionnement du cerveau humain.
- Apprentissage non supervisé : pas d'étiquetage de la donnée

6.1.10 Groupe TF1 : Philippe Toublant : Directeur Adjoint Data

Date : 07/08/2018

Type d'interview : Face à face

Identité de la personne l'interviewée

Nom : Toublant

Prénom : Philippe

Titre : Directeur Adjoint Data, Groupe TF1

- 1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*
 - Directeur Adjoint Data, Groupe TF1 depuis 3 ans
 - Statisticien de formation
 - Collecte et gère de la donnée
 - Produit des algorithmes afin d'avoir des indicateurs pour
 - ➔ Faire de la publicité ciblée
 - ➔ Faire de la recommandation de contenu
 - ➔ Améliorer la connaissance du client
 - Mise en place d'une nouvelle équipe

- 2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud ?*
 - Le cloud est de « l'informatique libre-service »

- 3) *Quelle est votre expérience du cloud ? Depuis combien de temps l'utilisez-vous à titre professionnel ?*
 - Infrastructure 100% cloud dès le début
 - ➔ « Data Management Platform » : Adobe Audience Manager en SaaS
 - ➔ Datalake chez Microsoft Azure
 - ➔ Quelques données chez AWS
 - ➔ « Dataviz » via Tableau en SaaS
 - ➔ Usage de cluster Hadoop en offre managée (Azure HD Insight)
 - L'usage du cloud était une évidence à la création de l'équipe
 - ➔ Rapidité de mise en œuvre
 - ➔ Exploitation des technologies du moment
 - ➔ Permet d'attirer des talents (attractivité de la plateforme)
 - ➔ Facilite l'échange de données
 - ➔ Facilite le travail à distance

4) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA*

- On met beaucoup de choses derrière l'IA, mais le plus souvent il s'agit de modèles de prévision parfois anciens, mais revisités.
- Présence de l'OpenSource très forte dans l'IA
- « Machine Learning » : l'algorithme s'écrit sous forme de formule
- « Deep Learning » : l'algorithme se construit lui-même

5) *Utilisez-vous ou prévoyez-vous d'utiliser une offre d'IA dans le cadre de votre activité professionnelle ?*

- Usage d'algorithme Spark sur plateforme Hadoop
- Certaines briques d'analyses cloud : Text to Speech

6) *Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ? Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?*

- Le cloud est obligatoire pour exploiter des solutions d'IA
- Il permet d'offrir la puissance de calcul et le stockage nécessaire aux algorithmes d'IA

7) *Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel (et du cloud) ?*

- Contrôle de l'appartenance de la donnée : lorsqu'on utilise des solutions d'IA clefs en main, il faut se poser la question : « Qui est le propriétaire final de la donnée ? » (Point important sur des solutions SaaS)
- Sécurisation des accès. Certaines données peuvent être soumises à des obligations légales (RGPD) ou exiger des contraintes de confidentialité. Il faut donc faire attention aux solutions retenues (ex : sécurisation du data lake via obligation d'y accéder à partir d'un réseau bien défini)
- Il faut parfois accepter que l'apprentissage réalisé sur sa propre donnée soit réutilisé par le fournisseur de la solution et bénéficie ainsi à ses autres clients.

8) *Quelle va être l'évolution de l'IA/cloud pour vous ?*

- Fiabilisation de l'IA
- Meilleure maîtrise du risque : précision accrue du machine Learning
- Besoins grandissant en IA
- Fusion des technologies avec l'être humain (Transhumanisme)

6.1.11 Microsoft : Clement Le Roux : Cloud Solution Architect AI & Data

Date : 27/09/2018

Identité de la personne l'interviewée

Nom : Le Roux

Prénom : Clement

Titre : Cloud Solution Architect

Type d'interview : document rempli par la personne

- 1) *Bonjour, pouvez-vous vous présenter, nous donner quelques informations sur votre parcours et votre position actuelle (rôles et missions) ?*

Cloud Solution Architect – en charge de l'adoption de la plateforme Azure chez les grands comptes français (CAC40)

- 2) *Que représente le cloud pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre cloud*

Le Cloud regroupe l'ensemble des services informatiques qui vont être fournis aux grandes entreprises pendant les 10 prochaines années. Les caractéristiques du Cloud sont l'élasticité, la scalabilité, la facturation à l'usage et l'utilisation à la demande.

- 3) *Que représente l'IA pour vous ? Pouvez-vous nous donner votre définition ? Quelles sont pour vous les caractéristiques d'une offre proposant de l'IA*

L'intelligence artificielle est aujourd'hui principalement mise en œuvre par des méthodes dites de Machine Learning. Le Machine Learning est caractérisé par l'utilisation de Dataset qui va servir de base d'entraînement pour un modèle prédictif. Une fois entraîné, le modèle prédictif est en mesure de réaliser des prédictions sur de nouvelles données. Une prédiction peut être la classification d'une donnée dans une catégorie, une valeur réelle, une recommandation, une détection d'anomalie dans le cas d'apprentissage supervisé ou un regroupement dans un cluster dans le cas d'apprentissage non supervisé. Un dernier cas d'application récente est le Reinforcement Learning, qui permet d'entraîner des modèles sur des problématiques beaucoup plus complexes comme gaming (Go, Pacman) ou véhicule autonome.

4) Êtes-vous fournisseur de solutions d'IA

- a. Si oui, quel est votre positionnement ? pouvez-vous nous faire une description de votre offre (marché, cible, missions, clients)

Oui, toutes les industries sont concernées, car l'IA s'appuie sur la valorisation de données, et toutes les entreprises, petites, moyennes ou grandes, quelle que soit l'industrie possèdent de la donnée qui peut être exploitée pour entraîner un modèle prédictif. La mission de Microsoft sur l'IA consiste à démocratiser son usage, au service de l'amélioration/augmentation des capacités humaines. Pour cela, différents services sont proposés pour simplifier l'intégration de l'IA dans les applications et le processus métier des entreprises.

- b. Depuis combien de temps votre offre (ou vos offres) est-elle (sont-elles) disponible(s) ?

Environ 3 ans.

- c. Si oui, vos solutions reposent-elles sur une architecture cloud si oui pourquoi ? Si non pourquoi ?

Oui, les différentes solutions sont proposées depuis la plateforme Cloud Azure. De ce fait, elles disposent des mêmes caractéristiques que le Cloud (élasticité, scalabilité, facturation à l'usage, on-demand). À noter cependant que dans un monde de mobilité, une trajectoire importante consiste également à pouvoir exporter des parties de l'IA directement sur les terminaux en mobilité voire privé de connectivité – c'est le concept d'IA « at the edge »

- d. Votre offre va-t-elle évoluer ? Dans quelles directions ?

Très certainement, les roadmaps d'évolution dépendent des groupes produits définissant la stratégie des produits. Un invariant étant la démocratisation des usages, la tendance sera probablement l'ouverture de solutions sur étagère type Cognitive Services extrêmement simple, vers la possibilité d'un peu plus de customisation des IA.

- e. Avez-vous des concurrents ? Si oui lesquels ?

Les grands acteurs du marché du Cloud : AWS + Google

f. L'IA est-il un enjeu important pour vous ?

Oui, car l'IA permet de créer de la nouvelle valeur sur un patrimoine existant (la data)

5) Pensez-vous qu'il y'a un lien entre l'IA et le cloud ? Si oui lequel ?

Oui, car les besoins en stockage et puissance de calcul que des IA performantes nécessitent ne sont disponibles que dans ce type de services.

6) Pensez-vous que le cloud est facilitateur du développement/de l'usage des solutions d'IA ?

Oui, par nature des caractéristiques du Cloud d'une part, d'autre part par l'ambition de démocratisation de son usage qui représente un enjeu business important en tant que techno créant de la valeur.

7) Pensez-vous que l'IA va transformer des métiers ? Si oui lesquels et de quelles façons ?

Oui très certainement, en augmentant les informations disponibles lors de la prise de décision (exemple dans le médical), en automatisant certaines tâches (exemple contrôle de qualité), en créant de nouvelles manières d'inventer (exemple en générant du contenu)

8) De façon globale, pensez-vous que l'IA est déjà mise en application en France dans le monde professionnel ?

Non, la démocratisation de l'usage de l'IA est un chantier en cours de démarrage. La prise de conscience de son potentiel et la capacité des entreprises à l'intégrer en se formant et en maîtrisant son usage vont prendre du temps.

9) Quels seraient pour vous les freins à l'adoption de l'IA dans le monde professionnel ?

La mise en œuvre d'IA custom, nécessitant des compétences avancées en Data Science.

L'incompréhension ou la non-maîtrise de la technologie.

Les problématiques éthiques et sécurités liées à l'impact de cette technologie.

10) Pensez-vous que l'IA a un intérêt/impact particulier pour les entreprises financières. Si oui pouvez-vous développer ?

Probablement, pas spécialiste du sujet. Les scénarios tournent probablement autour de la détection d'anomalie et du passage d'ordre sur les plateformes de trading.

11) *Quelle va être l'évolution du cloud pour vous ? (Nouvelles technologies, nouveaux défis.)*

Nouveaux moyens de stockage (exemple : ADN) et de calcul (Quantum Computing)

La sécurité va demeurer un enjeu majeur pour les plateformes Hyper Scale qui consolident des millions de serveurs.

12) *Quelle va être l'évolution de l'IA pour vous ?*

L'impact de l'IA dans la robotique est probablement à suivre

L'utilisation de plusieurs IA simultanées pour dépasser les limites de l'intelligence « monotâche »

Des IA autonomes capables d'apprendre de nouvelles choses par elles-mêmes, sans besoin d'entraînement à l'initiative des humains

6.2 AUTRES SOURCES

Intelligence Artificielle: Manuel de survie de Jean-Claude Heudin aux Editions Science Ebook

L'intelligence artificielle expliquée de Thomas Cambrai (autopublication)

Comprendre le deep learning de Jean-Claude Heudin aux Editions Science Ebook

La transformation digitale des sociétés de gestion en « SGP 3.0 » de l'association française de la gestion financière. (www.afg.asso.fr)

7 Glossaire

AI : Artificial Intelligence : Équivalent d' « Intelligence Artificielle » en anglais

API : Application Programming Interface : permet d'interagir de manière simple avec une application pour lui soumettre une demande et récupérer un résultat

BigData : Terme anglophone regroupant les technologies et les problématiques liés au traitement des très gros volumes de données.

Data scientist : Métier dont le rôle est d'analyser, exploiter et surtout donner du sens à des données de plus en plus nombreuses.

Fintech : startups qui utilisent les nouvelles technologies à destination de la finance

Framework : Ensemble d'outils et de fonctions informatiques formant un environnement de travail spécialisé pour la conception de logiciel. Il simplifie le travail des développeurs informatiques.

Natural Language Processing : Traitement automatique du langage naturel : regroupe les outils informatiques visant à interpréter et comprendre le langage humain.

OPCVM : terme financier pour « Organisme de Placement Collectif en Valeurs Mobilières ».

Processor : Processeur informatique : Puce électronique dédiée aux traitements informatiques. Elle peut être spécialisée dans un type de calcul (GPU (calcul graphique), TPU/NPU (machine learning)) ou généraliste (CPU)

Providers : Traduction de fournisseur en anglais. Terme que l'on retrouve souvent pour parler des entreprises proposant des services de cloud computing.

Python : Langage de programmation particulièrement utilisé pour exploiter certains algorithmes mathématiques ou automatiser des tâches simples.

Open Source : Un logiciel est dit « Open Source » lorsque le code informatique avec lequel il a été construit est accessible publiquement et gratuitement (ce qui n'implique pas forcément que le logiciel soit gratuit).

Ramdom Forest : (forêt d'arbres décisionnels en français) : Algorithme mathématique particulièrement performant pour résoudre des problématiques de prédiction.

Speech to text : Outils informatiques permettant à un ordinateur, smartphone ou tablette de retranscrire à l'écrit des phrases dictées par un humain.

Text to speech : Outils informatiques permettant à un ordinateur, smartphone ou tablette d'énoncer à l'oral le contenu d'un texte.

VL : Valeur liquidative. Prix d'une part dans un produit financier.

VM : Virtual Machine ou Machine Virtuelle : serveur informatique « virtualisé » c'est-à-dire installé au sein d'un serveur physique. Plusieurs serveurs virtualisés peuvent être installés sur un seul serveur physique permettant ainsi de proposer des environnements d'exécution multiples et isolés les uns des autres, tout en optimisant l'usage des équipements physiques.