



Opinions Libres

le blog d'Olivier Ezratty

Les avancées de l'intelligence artificielle – 6

Après avoir fait le tour des stratégies d'IA de quelques grands acteurs du numérique dont Google, IBM, Microsoft et Facebook et à leurs acquisitions, revenons aux startups du secteur en nous intéressant aux françaises.

Il est clair que l'IA est l'une des technologies clés du numérique, aujourd'hui et demain. Donc, au lieu de chercher à créer un Google, un Facebook ou un système d'exploitation français, il serait bon de s'intéresser à ce domaine prometteur, surtout dans la mesure où les plateformes correspondantes sont encore en devenir.

La recherche en IA en France

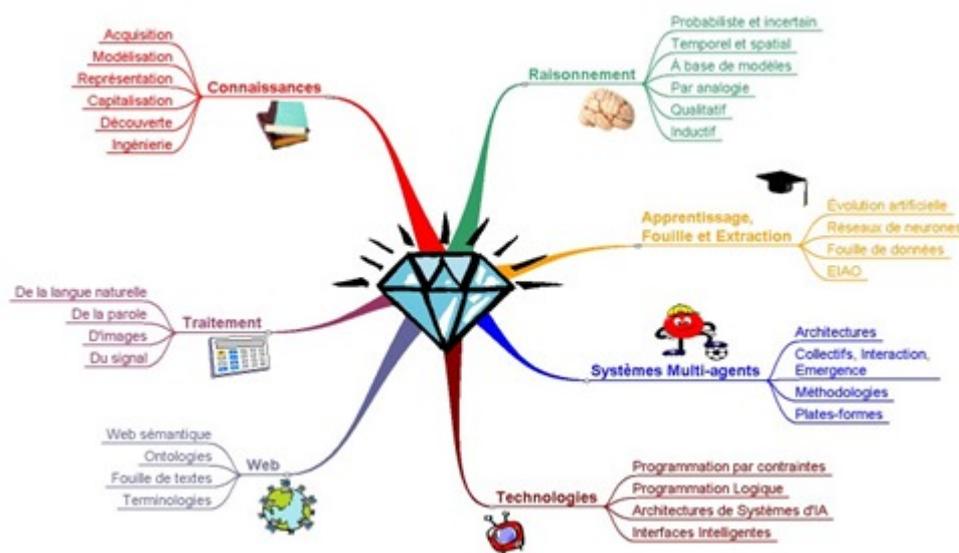
La recherche en IA est disséminée dans plusieurs laboratoires et dans des projets collaboratifs associant laboratoires publics et universités. Les deux premiers organismes se focalisant sur l'IA sont l'INRIA et le CNRS.

Que fait l'INRIA ? Un grand nombre des projets de recherche fondamentale en IA référencés sur leur site font appel aux techniques de l'IA, même s'ils ne sont pas forcément labellisés IA / machine learning / réseaux neuronaux. C'est ainsi le cas du projet **Orpailleur** mené à Nancy et dédié à la représentation des connaissances et au raisonnement. L'équipe planche sur l'extraction de données dans les bases de connaissances non structurées, et notamment dans le domaine de la santé, le même que celui qui est investi par IBM Watson et plein de startups. Ils collaborent notamment avec le centre de lutte contre le cancer de Nancy. L'équipe **Magnet** travaille quand à elle directement sur le machine learning et l'auto-apprentissage.

Les chercheurs français **se plaignent** en tout cas d'être délaissés en France dans la discipline. Ils ne sont certainement pas les seuls, au sens où de nombreuses disciplines se sentent délaissées dans la recherche publique.

Une association créée en 1993 fait la promotion de la recherche en IA, l'**AFIA**. Elle organisait en octobre 2014 une **conférence** de promotion de l'IA dans la recherche. On y identifie par exemple **Andreas Herzig** (IRIT, CNRS, Toulouse) qui travaille sur la modélisation de la logique et du raisonnement, **Hélène Fargier** (IRIT, CNRS, Toulouse) qui travaille notamment sur la programmation par contraintes, **Jérôme Euzenat** (LIG, Inria) qui planche sur la représentation et l'échange de connaissances et **Leila Amgoud** (IRIT, CNRS) qui est spécialisée dans la modélisation de l'argumentation.

Le défi pour ces chercheurs et leurs autorités de tutelle est de trouver des applications marchés de leurs travaux. En consultant la liste des participations d'**IT-Translation** qui est l'un des principaux financeurs de projets issus de l'INRIA, on constate que l'IA est souvent en filigrane de ces projets, mais pas forcément au niveau "plateforme" ou "couches de base".



Dans “**Economic Report of The President**”, le rapport annuel 2016 sur l’économie de la Maison Blanche, j’ai découvert deux données intéressantes : aux USA, en 2013, les startups ont créé 2 millions d’emploi et les entreprises traditionnelles 8 millions. Donc 20% ! Une proportion énorme sachant que dans le même temps, l’économie française a plutôt détruit des emplois et les startups n’en ont probablement créé que quelques milliers. Et surtout : la moitié de la R&D fédérale est dédiée à la défense ! Et au milieu des années Reagan, elle en représentait les deux tiers ! Cela explique pourquoi tant de projets autour de l’IA sont financés par la DARPA. Y compris trois défis lancés en 2004, 2005 et 2007 sur la conduite automatique, qui ont dynamisé les équipes de recherche de nombreuses universités sur le sujet. Nombre de ces équipes ont été ensuite recrutées par Google pour ses différents projets de voitures automatiques.

En France, la recherche dans l’IA semble mieux financée côté civil, même s’il est difficile de le vérifier par les chiffres. On ne s’en plaindra pas. A ceci près que la R&D militaire US a une qualité : elle est orientée vers des objectifs pratiques selon des cahiers des charges. De son côté, la recherche civile française fonctionne plutôt de manière très décentralisée et sans objectifs pratiques clairs, sauf lorsqu’elle est financée par des entreprises privées, surtout depuis la loi Pécresse de 2007. A méditer !

Startups “horizontales”

Voici les startups que j’ai pu repérer dans les solutions techniques d’IA plus ou moins génériques. Le champ de la reconnaissance audio est faiblement couvert par les startups françaises. Dans celle des images, on en a eu quelques cas anciens comme LTU qui a été racheté par le japonais JASTEC en 2005.

Il subsiste quelques acteurs spécialisés dans la recherche et qui ont intégré petit à petit des techniques d’IA dans leurs offres. Antidot et Sinequa sont anciens dans le paysage mais, à l’instar de nombreux éditeurs b2b, ils peinent à croître pour atteindre la taille critique, même s’ils commencent à se développer à l’international comme Sinequa qui y réalise maintenant 50% de son chiffre d’affaire.

J’indique comme dans l’article précédent entre parenthèse l’année de création et les montants levés lorsqu’ils sont disponibles. J’aimerais bien ajouter un troisième indicateur : le chiffre d’affaire, mais il n’est généralement pas disponible.

Antidot (1999, \$3,5m) est connu pour son moteur de recherche pour entreprises. Il propose une fonction de classification automatique de contenus ainsi que d’amélioration de la pertinence des résultats de recherche s’appuyant sur du machine learning.

Sinequa (2002, \$5,33m) est un fournisseur de solutions de big data et d'analyse de données pour les grandes entreprises. Il fournit un moteur de recherche sémantique capable d'exploiter les données issues de nombreux progiciels (ERP, CRM, gestionnaires de contenus, etc). La société a annoncé en 2015 investir dans le machine learning pour améliorer la performance de ses solutions.

Dataiku (2013, \$3,5m) fait évoluer les concepts de business intelligence et de data mining avec son Data Science Studio, un ensemble d'outils d'analyse de données qui exploitent du machine learning pour la création de modèles de données et de simulations.

Heuritech (2013) propose sa solution logicielle Hakken d'analyse sémantique, de tagging et classement automatiques de textes, images et vidéos sous forme d'APIs. Ils proposent aussi HeuritechDIP qui permet d'améliorer sa connaissance des clients et d'anticiper leurs besoins, évidemment, surtout dans les applications de commerce en ligne. Le tout s'appuie sur force marchine et deep learning. La startup s'appuie sur les travaux de recherche de deux laboratoires publics le CNRS LIP6 and l'ISIR de l'UPMC (Paris VI).



Les solutions de reconnaissance de visage qui évaluent l'âge sont généralement à côté de la plaque. L'habit (du visage) ne fait pas le moine ! Ici, sur le stand de Smart Me Up au CES 2016 de Las Vegas.

Proxem (2007, 1m€) propose une solution de traitement automatique du langage permettant de filtrer, analyser, tagger et classer automatiquement de gros volumes de données textuels, comme dans les commentaires d'utilisateurs dans les réseaux sociaux ou sites de e-commerce. Le tout s'appuie sur des techniques de machine learning et de deep learning. L'outil permet notamment d'explorer les données analysées de manière visuelle pour identifier des patterns et signaux faibles.

Smart Me Up (2012, 3m€), vu aux CES 2015 et 2016 propose une solution logicielle d'analyse des visages. Elle détecte l'âge, le comportement et les émotions des utilisateurs. La solution est bien entendu plutôt commercialisée sous forme de brique logicielle en marque blanche utilisable dans des applications métier.

Regaind (2014, 400K€) propose une solution de tri automatique de photos en cloud s'appuyant sur du machine learning et de deep learning. Elle permet de trier les photos sous un angle à la fois narratif et descriptif et de les tagger automatiquement.

Moodstocks (2008) propose une solution mobile de reconnaissance d'images, fournie sous la forme d'APIs et d'un SDK multi-plateforme (la startup a été acquise par Google en juillet 2016).

beaucoup sur des solutions open source du marché. Ils travaillent déjà avec la SNCF, mais pas sur des robots.

Ubiant (2011), basé à Lyon, était également présent au CES de Las Vegas en 2015 et 2016. Il propose une solution matérielle et logicielle de gestion de la maison intelligente, de l'éclairage et de l'énergie qui s'appuie sur du machine learning et sur le Luminion (*ci-dessous*) un objet connecté interagissant avec l'utilisateur via des LED de couleur indiquant si la consommation du foyer est supérieure à celle du voisinage. C'est une offre b2c.



Vivoka a développé Lola, un logiciel de contrôle des équipements de la maison connectée. Elle s'appuie sur une box reliée à Internet qui se pilote via une application mobile et par commande vocale. Le projet lancé sur Kickstarter n'a pas porté ses fruits.

Iqspot (300K€) est une startup bordelaise qui analyse la consommation énergétique des bâtiments et sensibilise ses occupants pour la diminuer. Le tout avec du machine learning. C'est une participation de IT-Translation.

Xbrain (2012) est une startup française établie dans la Silicon Valley ainsi qu'à Paris et Lille qui se spécialise dans les applications de l'IA à l'automobile et la robotique. Sa plateforme xBrain Personal Assistant permet de créer des agents conversationnels. Elle s'appuie sur la reconnaissance vocale, sur la gestion de contexte, sur la détection des intentions et la gestion de règles. Son créateur, Gregory Renard, planche sur l'IA depuis près de 20 ans.

Scortex (2016) développe des solutions matérielles et logicielles apportant l'autonomie aux robots et objets connectés qui intègre notamment la reconnaissance d'images et de la parole. Ils ont même développé un chipset à base de réseaux neuronaux.

Commerce et marketing

L'écosystème français a toujours été prolifique en startups b2b et b2c dans le secteur du e-commerce et du marketing. Il est donc normal d'y retrouver quelques startups intégrant de l'IA.

AntVoice (2011, \$3,5m) propose une solution de recommandation prédictive pour les sites de e-commerce qui s'appuie sur de l'intelligence artificielle. C'est un spécialiste du big data marketing. La solution analyse la pondération de la relation entre internautes et produits et s'appuie sur la théorie des graphes.

Datapred (2014) propose également une solution d'analyse prédictive basée sur du machine learning. La société cible divers marchés professionnels dont celui de la distribution, en plus de la finance, de la logistique et

de la santé. Elle permet par exemple de simuler des hypothèses marketing et leur impact sur une chaîne logistique de distribution en tenant compte d'un grand nombre de paramètres. Comme c'est souvent le cas, le lancement d'un projet requiert une bonne part de service et de personnalisation avant sa mise en oeuvre opérationnelle.

DataPublica / C-Radar (2011) est une société qui propose une solution en cloud de marketing prédictif B2B permettant de cibler les bons prospects. Elle s'appuie sur l'exploitation des données administratives et financières des entreprises issues de sources publiques, des sites web associés, des réseaux sociaux et des mentions dans les médias. Ces données permettent alors de segmenter automatiquement les clients, de priorisation de ces segments, le tout s'appuyant sur un apprentissage supervisé. L'approche permet par exemple de segmenter les startups d'un secteur d'activité donné (Medtech, Fintech). La société est une autre participation d'IT Translation.

D'autres startups françaises se positionnent sur ce créneau comme **Compellia** (2015), qui analyse des sources données ouvertes et identifie des événements clés de la vie des entreprises pour créer des listes de prospects qualifiés, sachant que le processus est spécifique à chaque marché.

Il y a aussi **TinyClues** (2010, \$7,37), une startup plus établie qui utilise des solutions de machine learning pour identifier les produits que les clients de sites de vente en ligne sont le plus susceptibles d'acheter, histoire d'optimiser les campagnes marketing ciblées au niveau du ciblage comme des messages et des offres.

Search'XPR (2013, \$3,2m) est une startup créée à Clermont-Ferrand qui a créé le concept de "sérendipité psycho-cognitive" issu d'une thèse soutenue en 2010 par Jean-Luc Marini, l'un des cofondateurs de la société. Le concept est mis en oeuvre dans la solution Oorace, destinée au commerce en ligne et même traditionnel. Elle permet d'analyser l'état d'esprit du consommateur et d'évaluer sa réceptivité à des propositions commerciales inattendues, affichables notamment dans des offres ciblées s'apparentant à du "retargeting publicitaire" un peu moins bourrin que celui de Criteo. Le tout s'appuie sur de l'analyse syntaxique des sites visités et du parcours du visiteur, associant algorithmes et sciences cognitives analysant les "émotions" des utilisateurs, avec à la clé une **augmentation** des taux d'achat et du niveau des paniers moyens. Le service est fourni sous la forme d'APIs en cloud. Reste à savoir si les algorithmes relèvent réellement de l'IA et comment ils fonctionnent. C'est la "secret sauce" de la société, vaguement **documentée ici**. Pas forcément de l'IA au sens classique du terme, mais plutôt une algorithmie bien sentie, probablement astucieuse dans sa forme, qui permet d'éviter la force brute de nombreux solutions de machine learning.

Dictanova (2011, 1,2m€) est une société nantaise à l'origine d'une solution d'analyse textuelle des feedbacks clients dans les réseaux sociaux ou sites de vente en ligne, en liaison avec les outils de CRM pour optimiser la relation client. Les techniques utilisées comprennent l'analyse sémantique de textes et la classification automatique. La solution est fournie en cloud. C'est une autre participation de IT-Translation.

Modizy (2012, \$275K) propose un assistant d'achats dans la mode basé sur un algorithme d'intelligence artificielle. Modizy propose aussi une place de marché reliant consommateurs et marques.

Do You Dream Up (2009) propose une solution de chat automatique pour les sites en ligne. La société est basée à Paris, Bordeaux et Londres. Et elle a déjà une bonne douzaine de clients grands comptes ayant déployé sa solution.

Tastehit (2014) utilise du machine learning et du big data pour personnaliser les sites de e-commerce en temps réel. Donc, une offre b2b.

CompareAgences (2012) intermédie la relation entre agents immobiliers et particulier dans le cadre de la vente de biens. La startup emploie 12 personnes et génère 200 000 visiteurs uniques par mois. 1000 agences

immobilières sont intégrées en France. Le tout est à base de machine learning, sans plus de précisions.

Cypheme (2015) est une startup proposant une application mobile de détection de produits contrefaits, s'appuyant sur un algorithme de machine learning appliqué à la qualification d'images. La startup est accélérée chez Microsoft Ventures à Paris. C'est une sorte de Shazam de la contrefaçon.

Santé

C'est un domaine très porteur pour les applications de l'IA. Seulement, voilà, nous sommes un peu à la traîne dans l'une de ses grandes applications : la génomique. Mais la santé va au-delà de la génomique, heureusement.

CardioLogs Technologies (2014) a créé une solution d'interprétation automatique des électrocardiogrammes (ECG) en temps réel s'appuyant sur du machine learning. Uberisation en puissance des cardiologues ? Pas si vite ! Cela permet surtout de rendre un suivi plus régulier des patients à risques ou atteints de maladies chroniques.

DreamUp Vision (2015) est une startup issue de Dreamquark, une startup spécialisée dans l'analyse de données pour la santé et les assurances. Elle propose une solution d'analyse des images de la rétine obtenues par un fond de l'œil traditionnel. Elle permet de détecter les rétinopathies diabétiques émergentes aussi bien que les ophtalmos. Elle se situe dans un mouvement comprenant quelques autres acteurs dans le monde qui traitent automatiquement les résultats d'imagerie médicale. C'est ainsi le cas d'une autre startup francilienne, **Qynapse** qui analyse de manière itérative les résultats d'IRM cérébrales pour suivre l'évolution de traitements, notamment dans la lutte contre les cancers du cerveau.

Dexstr.io (2014) est une startup toulousaine fournissant la solution Inquiro qui exploite les données médicales non structurées pour faciliter la recherche d'informations pour les sociétés de pharmacie. En gros, c'est de la recherche documentaire, un peu comme le font Sinequa et Antidot, mais avec un tuning adapté à la documentation scientifique dans la santé. Leur concurrent serait plutôt l'application d'IBM Watson à l'oncologie. C'est encore une participation de IT-Translation.

Khresterion (2014) propose un logiciel d'aide au diagnostic et à la prescription pour le et les cancers. La solution fonctionne sur un principe voisin de celui d'IBM Watson, compulsant la littérature scientifique et les données des patients pour proposer divers traitements avec leurs avantages et inconvénients comme les effets secondaires. La société aurait comme prescripteur des organismes de remboursement comme Humanis, Axa et la Maaf. Sa solution commence aussi à être utilisée dans la finance, là où les cycles de vente sont probablement plus courts.

Industrie

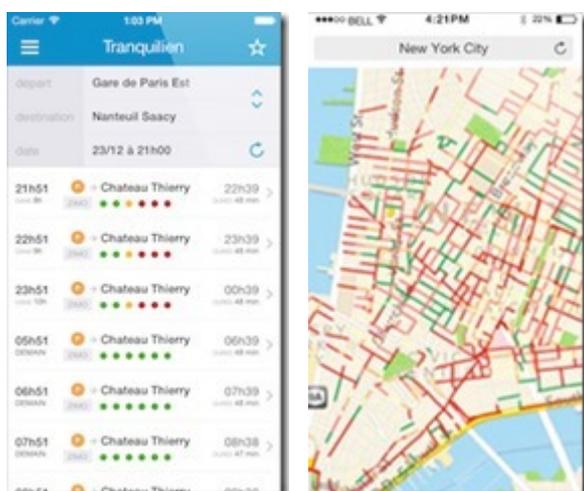
Il existe probablement de nombreux acteurs dans ce vaste domaine où le machine learning peut avoir plein d'usages.

Citons par exemple le cas de **DCbrain**, issu de Telecom Paristech, et spécialisé dans la modélisation de flux physiques (eau, gaz, électricité, ...) pour les utilities, qui s'appuie sur des solutions de big data et de machine learning pour identifier des tendances et signaux faibles. Cela permet de piloter automatiquement les réseaux, de faire de la maintenance prédictive et de modéliser le fonctionnement des réseaux.

Applications métiers

C'est là que la créativité est la plus développée, comme nous l'avons vu dans l'article précédent de la série au sujet des startups américaines.

Snips.ai (2013, \$6,3m) est une startup connue du secteur de l'IA créée par Rand Hindi (prix du MIT30 en 2015), Mael Primet et Michael Fester. Leur dernière levée de fonds de 5,7m€ en juin 2015 présente la particularité d'associer Bpifrance avec des investisseurs américains, en plus de business angels tels que Brent Hoberman et Xavier Niel. L'équipe comprend 35 personnes : des data-scientists, des développeurs, designers et quelques marketeurs. Leur positionnement est large et un peu vague : rendre la technologie invisible et les usages intuitifs via de l'IA. A ce titre, la startup a développé des applications expérimentales telles que :snips (un ensemble d'applications de recherche pour iOS dont un clavier virtuel intelligent pour la recherche d'adresses), Tranquilien (qui prédit les places disponibles dans les trains de banlieue), Parkr (la même chose pour prédire les places de parking), Flux (qui identifie le trafic mobile en s'appuyant sur les données des smartphones), RiskContext et SafeSignal (identification de risques d'accidents sur la route). La startup planche aussi sur des applications verticales : pour les véhicules connectés, dans l'hôtellerie, la maison connectée et les loisirs numériques. Le tout s'appuie sur force machine et deep learning, modèles probabilistiques, traitement du langage, gestion de graphes et aussi encryption des données pour garantir la vie privée. Derrière la vision, l'implémentation et l'expérimentation, on leur souhaite de réussir la businessmodelation.



Jam (1m€) a créé un agent conversationnel SMS pour étudiants. Ils ont ISAI Ventures dans leur capital. La solution utilise une combinaison d'IA et de vrais intervenants humains pour assurer une bonne qualité des réponses. Leurs outils d'IA sont en open source.

Julie Desk (2014, \$993K), basé à Paris, propose un service d'assistante virtuelle fonctionnant sous la forme d'un agent conversationnel opérant en français et en anglais. Il gère surtout votre agenda et réponds à vos mails à votre place pour prendre des rendez-vous avec vos interlocuteurs. Comme pour Jam, l'agent fonctionne en mode supervisé par des opérateurs ce qui permet d'assurer une bonne qualité de service. Les tarifs vont de 50€ à 80€ par mois. Il est notamment utilisé par des entrepreneurs de startups. Mais l'agent ne répond pas encore au téléphone.



Riminder (2015) est une startup spécialisée dans les RH qui s'appuie sur du deep learning pour proposer des outils d'aide à la décision. Il aide les chercheurs d'emploi à construire leur parcours professionnel et les actifs à développer leur carrière, en exploitant une base de connaissance de plusieurs millions de parcours de cadres.

White (2015) est une startup hébergée à l'accélérateur de Microsoft Ventures à Paris qui permet la saisie automatique de pièces comptables pour l'expertise comptable et l'audit. L'outil est capable de comprendre la structure du document et de le traiter convenablement dans son environnement. Il va au-delà des solutions traditionnelles d'OCR (optical characters recognition).

niland (2013) est une autre participation de IT-Translation, la structure de valorisation des projets de recherche issus notamment de l'INRIA. Mais la startup a été créée par des anciens de l'IRCAM et s'appuie sur 10 années de travaux de recherche. Elle utilise le deep learning analysant le contenu de la musique pour rendre son exploration dans les plateformes de diffusion plus intelligente. Elle identifie les similarités entre morceaux pour les classer automatiquement. La solution sera exploitée par CueSongs (UK, une société fondée par le chanteur Peter Gabriel) et motionelements (Singapour) qui sont dédiés aux professionnels de la musique. La solution est aussi illustrée par le service en ligne www.scarlett.fm et s'appuie sur Soundcloud pour vous permettre de créer une web radio personnalisée en fonction de vos goûts.

Yseop (2008) propose son agent conversationnel **Savvy**. Nous l'avons déjà évoqué dans le **troisième article** de cette série. La société propose également une solution de Business Intelligence qui est capable d'extraire intelligemment des données de bases structurées (ERP, CRM, etc) pour les convertir en synthèse en langage naturel, après moulinage dans un moteur de déductions, et qui plus est, dans plusieurs langues.

Dhatim (2008) automatise la gestion des factures et le contrôle des déclarations sociales avec comme premiers clients les opérateurs mobiles (pour les factures) et d'autres (pour les déclarations sociales). Dans ce dernier cas, la solution permet d'éviter de générer des incohérences dans les déclarations sociales et les pénalités qui vont avec les contrôles qui sont eux inévitables. La solution s'appuie sur une combinaison de centaines de règles métiers et de machine learning qui déclenche des actions automatisées.

Séline (2013), édité par la société Evi, propose une panoplie d'applications bureautiques intégrant un agent conversationnel permettant de dialoguer et poser des questions en langage naturel. On y trouve notamment un traitement de texte, un tableur, un gestionnaire d'agenda, un carnet d'adresses, un gestionnaire de tâches, une médiathèque, un logiciel de gestion de finances et un gestionnaire de messagerie instantanée. Dilemme classique : faut-il recréer tout un existant complexe pour y intégrer une nouvelle fonction ou ajouter cette fonction aux produits existants du marché (Microsoft Office, Open Office). Question d'ouverture, de simplicité de mise en oeuvre et de modèle économique!

Dans mon **précédent article** sur l'écosystème entrepreneurial de La Réunion, j'avais aussi identifié quelques startups qui utilisent le machine learning : **logiCells** (ERP sémantique) et **Teeo** (analyse de consommation d'énergie pour les entreprises). A contrario, certaines startups font appel à des briques d'IA comme le machine learning mais préfèrent ne pas l'évoquer dans leur communication.

Ce tour est probablement incomplet et les oubliés du secteur se feront inmanquablement connaître pour intégrer cette liste que je mettrai à jour au fil de l'eau. A vrai dire, d'ici peu de temps, l'usage de machine learning sera aussi courant dans les startups que l'appel à des bases de données NoSQL : une banalité !

Le top du top de la startup d'IA ? Utiliser l'IA dans une solution d'agent conversationnel en cloud qui fait du big data sur des données issues de l'IOT en sécurisant les transactions via des Blockchains. Le Bingo de la startup d'IA est lancé !

Dans le prochain épisode, nous **examinons le cerveau d'un peu plus près** pour comprendre la complexité de sa modélisation, puis, ensuite, nous étudions les **facteurs d'accélération et de ralentissement des progrès dans l'IA**.

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

Episode 1 : **sémantique et questions clés**

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

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

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

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

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

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

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

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

Cet article a été publié le 17 avril 2016 et édité en PDF le 18 mars 2024.
(cc) Olivier Ezratty – “Opinions Libres” – <https://www.oezratty.net>