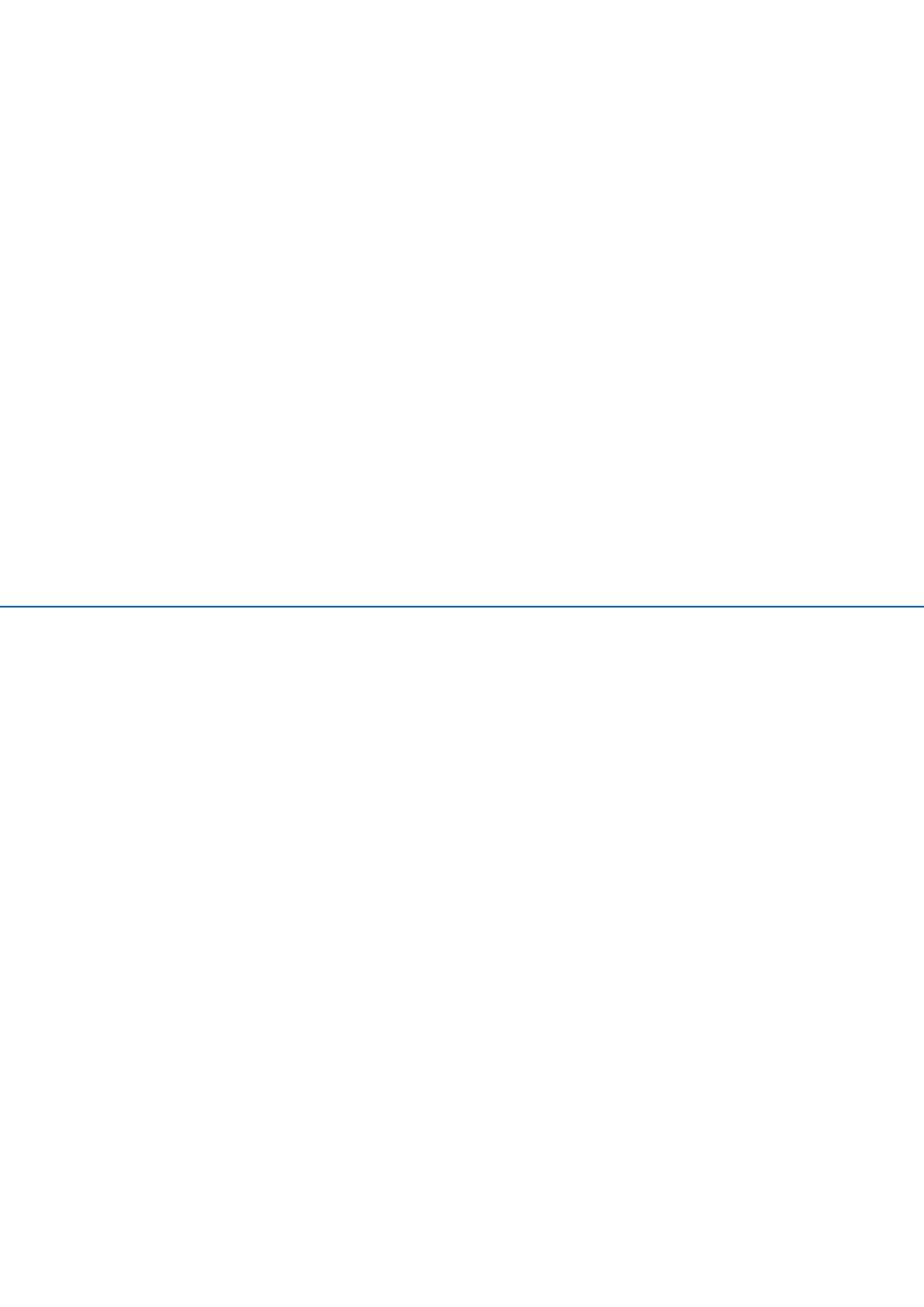


(Re)Launch,
Scale, & Lean AI
with Cloud



AI Cloud

(Re)Launch, Scale and Lean AI with Cloud



À propos d'Artefact,

Artefact incarne la prochaine génération d'agences digitales centrées sur l'utilisation de la data et l'intelligence artificielle. Animée par des ingénieurs et marketeurs aux profils uniques, l'agence propose des services complets en data, média, création et technologies (big data et intelligence artificielle) rehaussés d'une activité de conseil stratégique à forte valeur ajoutée. La société, dont le siège est à Paris, compte aujourd'hui plus de 1000 collaborateurs à travers le monde, dans 22 pays. Cette agence indépendante internationale fondée par trois anciens de l'Ecole Polytechnique : Vincent Luciani, Philippe Rolet et Guillaume de Roquemaurel a connu depuis sa création en 2015 un taux de croissance très soutenu avec un triplement de son chiffre d'affaires en 2016 (chiffres 2017). Le titre Artefact est coté sur le marché Euronext Growth Paris Stock Exchange.

artefact.com

Intro,

Chers Amis, Chers Humains, Chers Robots,

En créant Artefact nous avons voulu une agence moderne, pensée nativement pour l'ère du digital, de la data et de **l'intelligence artificielle***. Les GAFAX dominent en ce moment le classement des marques préférées des consommateurs et ce, grâce à la maîtrise de ces sciences. Ainsi, notre mission chez Artefact est d'allier d'un côté ces technologies, et de l'autre la créativité des marques.

Nous croyons en effet dur comme fer à l'alliance de l'homme et de la machine, de l'ingénieur et du créatif, et finalement au Chief Digital Officer augmenté. Artefact fait à ce titre collaborer à travers le monde, un millier de talents très différents : d'un côté des chercheurs, data scientists, ingénieurs, et de l'autre des experts en digital, des chefs de projets et des créatifs. Cela constitue au quotidien une formidable rencontre de cultures et de talents, et souvent une belle auberge espagnole.

C'est peu de le dire : l'intelligence artificielle génère des doutes, voire des peurs et apportera son lot de défis pour vous les entreprises :

– **Un premier défi est l'enjeu organisationnel.** Il va falloir inventer de nouveaux métiers, recruter des talents, former en continu, et assurer la collaboration transverse de nombreuses fonctions différentes de l'entreprise... Notre conviction est que l'intelligence artificielle est le nouveau digital.

Aujourd'hui, se sont formés des Chief Digital Officers dans les organisations, je suis convaincu que l'on verra apparaître demain des Chief AI Officers pour surmonter ces challenges.

– **Un second défi est la balance entre le recrutement des profils IA d'un côté et l'automatisation des tâches de l'autre.** Les économistes sont en désaccord sur le sujet et on ne sait pas de quel côté la balance va pencher. L'Histoire donne raison à un effet positif de l'innovation sur l'emploi : la période des trentes glorieuses a été la période de plus forte productivité en France au XXème siècle et cette période a également connu le plein emploi ! Quoi qu'il en soit, il va falloir trouver les talents formés à l'IA, et les attirer, et cela va être un challenge ! L'Europe devrait avoir besoin de 3M de data scientists d'ici 2020 et elle en forme aujourd'hui moins de 1000 par an ! Outre-Atlantique, les PhDs spécialisés dans ce domaine, fraîchement débauchés chez les GAFAX, sont rémunérés en moyenne entre 300,000 et 500,000 \$/an quelques années après l'école.

– **Le troisième défi est la formation :** aujourd'hui l'Éducation nationale se cherche sur ces sujets, et le nombre de formations disponibles ne répond pas à la demande. Il est donc clé que le monde de l'entreprise prenne les devants et mette en place ses programmes propres. Via le e-learning, les MOOCs et l'ensemble du contenu très largement accessible et souvent gratuit sur le sujet, c'est aujourd'hui beaucoup plus facile de se former. J'ajoute que ces programmes devront être remis à jour constamment, compte tenu des avancées quasi permanentes des outils, des méthodes.

– **Le dernier défi, peut être moins évident, c'est la transparence.** Plus les couches d'entraînement des modèles sont profondes, et moins les résultats sont interprétables. Il est très difficile de communiquer des recommandations, parfois contre-intuitives, de façon "boite noire" et donc il y aura un enjeu de communication, de transparence autour du **"retro-engineering"*** des algorithmes.

Aujourd'hui, lorsque l'on interroge nos clients, la grande majorité ont des projets IA lancés ou en cours de déploiement. L'enjeu est donc l'industrialisation. Et celle-ci devrait démarrer dès aujourd'hui ! Pourquoi ?

– **Les barrières technologiques historiques se lèvent** : les infrastructures sont désormais prêtes, on citera les progrès considérables du cloud en privacy et sécurité, la généralisation du recours aux **APIs*** permettant de requêter plus facilement des bases de données, la standardisation des méthodes de **calculs distribués*** permettant de paralléliser les calculs. Les données sont de mieux en mieux organisées, et sont souvent centralisées dans des **datalakes***. Donc finalement, ces barrières historiques sont levées ou en passe de l'être.

– **Ensuite, c'est une course contre la montre qui démarre.** N'en doutez pas : 100% des entreprises auront investi directement ou indirectement dans l'IA en 2020. C'est donc une course qui démarre pour l'utilisation de ces technologies, avec à la clé, des possibilités de récupérer des parts de marché, avec un premium pour les early adopters.



– **Pour finir, vous ne pourrez pas rattraper le train en marche.** Les avancées technologiques suivent une croissance exponentielle. A titre d'exemple, il y a eu autant d'évolutions technologiques sur l'ensemble du XX^{ème} siècle que de 2000 à 2018. Il est donc clé de construire des fondations aujourd'hui, pour ne pas être dépassé dans 3 ans.

Pourquoi un livre blanc sur le cloud et l'IA ?

Nous avons voulu, à travers cet ouvrage, apporter une réponse structurée sur les grands défis que traversent aujourd'hui les entreprises en affrontant ces problématiques. Nous abordons ces problématiques par un constat : le déluge de données et la récente mise à disposition d'infrastructures agiles ont permis l'essor dans l'entreprise de l'IA, c'est le "Launch". En revanche, si près de 90% des grandes entreprises évoquent avoir essayé l'IA, moins de 10% évoquent la possibilité d'avoir de l'impact. C'est le "Scale" qui nécessite la mise en place de fondations précises afin de réussir ses projets. Enfin, une fois l'IA démocratisée dans l'entreprise, elle pourra elle-même devenir un objet de simplification et de productivité : c'est le "Lean".

Au travers de ces trois étapes simples : **Launch, Scale, Lean**, nous avons voulu aider nos lecteurs, confirmés ou en recherche de repères, à construire leur plan de travail pour atteindre un impact mesurable et scalable dans toutes les organisations.

J'en profite pour remercier tous ceux sans qui ce livre n'aurait jamais vu le jour à commencer par ceux qui font Artefact tous les jours : nos collaborateurs et nos clients.

Nous vous souhaitons une excellente lecture !

Vincent Luciani.

Co-fondateur et Global COO d'Artefact



Merci !

Artefact tient à remercier particulièrement tous les contributeurs qui ont permis l'écriture et la réalisation de ce livre blanc. Merci à toutes les personnes qui ont répondu à nos questions et qui nous ont permis de lancer ce projet : Valérie Coscas (Amazon), Sébastien Wallet (Objenious), Michelle Lamberti et Thaïs Weber (Uber), Clément Cardi (Google), Rémi Cauchois (Orange), Steve Cheney (Allianz), Jean-Philippe Desbiolles (IBM), Kenza Ibnatty (Microsoft) et Pierre Fournier (ManoMano).

Un grand merci à toute l'équipe d'Artefact : Justine Nerce, Solène Marie, Pascal Coggia, Damaris Peillon, Elisabeth Gizard, Matthieu Myszak, Camille Le Gonidec, Camille Dugast, Benjamin Naert, David Molina, Cédric Ly, Guillaume Léger, Matthieu Montaigu, Sarah Bugalo, Hanania Ouazan, Dayvid Kayal, Douglas Willcocks, Diane Ly, Paul Essig, Benoît Lamairesse, Paul de Balincourt, Tristan Silhol, Jérôme Harousseau, Vincent Luciani, Guillaume de Roquemaurel, Philippe Rolet et Edouard de Mezerac pour tout leur travail, leur implication et leur aide précieuse sur ce projet !

Sommaire

01

(Re)launch AI

L'essor de l'IA dans le monde de l'entreprise : le lancement de multiples initiatives

p.14

Notre monde se dataifie à toute vitesse 18

- **La transformation digitale et l'adoption massive du machine-to-machine** 18
 - Homodataicus : l'Homme créateur de données 20
 - Le machine-to-machine explose grâce à l'IoT 20
 - Une valorisation du Big Data nécessaire mais encadrée 21
- **Un indispensable dépassement des limites physiques et intellectuelles** 22

90% d'adoption de l'IA via une pluralité d'initiatives 24

- **Le passage du Big Data à l'IA** 24
 - Une trajectoire scientifique, philosophique et culturelle 24
 - L'IA, comme imitation de la pensée humaine 30
 - Machine learning et deep learning : concrètement, qu'est-ce que c'est ? 31
 - La révolution neuronale, source d'intelligence 36
 - Un algorithme, ça ressemble à quoi ? 36
 - Demain, une super-intelligence ? 42
- **Au-delà des fantasmes, l'IA n'est pas magique** 43
 - Des performances dépendantes de la qualité d'alimentation 43
 - Un nécessaire "artisanat" de l'artificiel 43
 - L'IA augmente l'humain mais ne remplace pas son intelligence 44
- **L'IA, un levier à valeur convoité** 45
 - Les gouvernements font de l'IA une priorité économique 45
 - Les organisations non lucratives mettent en place l'AI for good 47
- **Les entreprises s'emparent de l'IA et lancent de multiples initiatives** 51
 - Maturité sectorielle : tous sortis de l'enfance 51
 - De multiples initiatives à valeur 51

Le cloud catalyse ces initiatives 62

- **Le cloud et l'IA sont-ils indissociables ?** 62
 - Le besoin d'abstraction 62
 - Le besoin de puissance des machines 62
 - Le besoin d'optimisation des coûts 64
 - Le besoin d'actualisation de l'IA 66
 - Le besoin de collaborativité 66
- **Entre cloud et IA : le positionnement stratégique des géants de la tech** 68
- **Quel choix d'infrastructure pour accueillir l'IA ?** 69
 - Cloud public, privé ou hybride : une adoption sur-mesure 69
 - Devenir agnostique des solutions cloud pour maximiser la valeur de l'IA 71
 - L'essor de la portabilité, un nouvel élan 71

02

Scale AI

Bâtir de solides fondations pour faire de l'impact à l'échelle

p.74

Libérer la donnée	79
<ul style="list-style-type: none"> – Construire un patrimoine data exploitable, permettre un accès rapide et flexible à la donnée 79 <ul style="list-style-type: none"> • Une donnée disponible : fournir un accès centralisé à des données décentralisées 79 • Une donnée compréhensible : fournir un accès centralisé à des connaissances décentralisées 80 • Une donnée exploitable : définir et lancer des stratégies de Quality-by-Design 80 – Garder l'humain au coeur de la transformation 82 <ul style="list-style-type: none"> • Créer et diffuser une culture data et IA 82 <ul style="list-style-type: none"> – Pour tous : Démystifier, "IA"alphabétiser" et Engager 82 – Pour les opérationnels : recentrer les efforts sur la créativité 84 – Pour les experts de la donnée et les équipes dirigeantes : faire évoluer les mentalités 84 <ul style="list-style-type: none"> • Apprendre à optimiser des boîtes noires 84 • Penser re-use et open source 84 • Penser cloud 84 • Recruter, développer et retenir les talents 85 <ul style="list-style-type: none"> – Recruter une équipe pluridisciplinaire 85 <ul style="list-style-type: none"> • Le mariage entre technique et métier au sein de l'équipe coeur 85 • Le soutien des experts dans l'équipe étendue 86 • Le sponsoring des change leaders 87 – Développer et retenir les talents 88 <ul style="list-style-type: none"> • Investir dans la formation 88 • Fidéliser : le défi des Chief People Officers 89 	
Installer les bons enablers pour accueillir l'IA au coeur de l'entreprise	92
<ul style="list-style-type: none"> – Mettre en place une infrastructure et des outils "AI friendly" : exploration facilitée, passage à l'échelle accéléré, rentabilité maximisée 92 <ul style="list-style-type: none"> • Les 4 partis pris d'Artefact sur la manière d'optimiser ses infrastructures pour passer le cap de l'industrialisation 92 <ul style="list-style-type: none"> – Simplifier vos infrastructures avec les microservices 92 – Packager vos "data products" grâce aux conteneurs 92 – Réunifier vos infrastructures cloud et on premise grâce à Kubernetes 93 – Paralléliser vos opérations de test et de production grâce à la "Livraison Continue" 94 • Pour accélérer la transformation, développer une intelligence artificielle "as-a-service" 96 <ul style="list-style-type: none"> – Utiliser la boîte à outils des solutions cloud 96 – Utiliser des outils collaboratifs de création et visualisation de workflows data 97 – Lancer des IA labs pour accélérer et créer massivement de la valeur 98 <ul style="list-style-type: none"> • Pourquoi mettre en place un IA lab ? 98 <ul style="list-style-type: none"> – Qu'est ce qu'un IA lab ? 98 – Les grands enjeux 98 • Les 5 convictions d'Artefact sur la mise en place d'un IA lab 99 <ul style="list-style-type: none"> – Choisir les bons cas d'usage 99 – S'organiser en feature teams 99 – Décomposer la complexité 99 – Assurer la montée en compétences 99 – Rendre scalable 99 • L'organisation de l'IA lab 100 • Le process de réalisation d'un cas d'usage 101 	

Développer une Intelligence Artificielle sécurisée et conforme	104
--	-----

- La protection des données, un enjeu clé 104
- Les 12 principales menaces associées à l'utilisation du cloud public 105
- Les 5 chantiers à lancer pour sécuriser ses infrastructures et respecter les données de ses clients 106

03

Lean AI

Optimiser la chaîne de production de l'IA pour maximiser la valeur

p.112

Optimiser la chaîne de production IA	116
--	-----

- **Les 7 "wastes" des projets d'Intelligence Artificielle** 116
 - Extra-processing : sur-délivrer par rapport aux attentes clients 117
 - Dépense d'énergie : des efforts humains ou machine qui n'apportent pas de valeur incrémentale 118
 - Inventaire : la création d'intelligence pas assez autonome 119
 - Sur-production : l'IA arrive trop tôt par rapport à la maturité de l'entreprise, ou de manière non structurée, non réfléchie 120
 - Attente : temps inoccupé, dans l'attente d'une livraison ou d'une décision managériale 121
 - Défauts : le produit IA ne répond pas aux besoins (dévie des attentes du client final) 122
 - Talents : échec à capitaliser sur les compétences et les connaissances internes 123
- **La réponse du Lean AI aux "7 wastes" des projets d'Intelligence Artificielle** 124
 - L'optimisation de la chaîne de production via le développement de data products 124
 - La standardisation : une contrainte culturelle, source des principales avancées technologiques .. 124
 - Les data products : des Lego intelligents 124
 - Une emphase forte sur le dernier kilomètre 125
 - La distribution des data products 125
 - L'IA Analytics 125
 - L'IA Product Owner ethnographe 126

Le plan d'action du CAIO	127
--------------------------------	-----

01

(Re)Launch AI

L'essor de l'IA dans le monde de l'entreprise : le lancement de multiples initiatives

01

“Launch” ou “Relaunch”, l’IA est une découverte perpétuelle pour les organisations, quelque soit leur niveau de maturité.

Pour celles qui portent leurs premières initiatives IA, les risques d’échec sont nombreux : manque d’impact, coût élevé, difficulté à lancer un prototype en production, etc. Ainsi, pour ne pas passer à côté de sa transformation IA, des étapes clés sont nécessaires, du changement de la culture d’entreprise aux impacts organisationnels profonds.

Intro.

Mêmes les entreprises les plus avancées auront besoin de fréquemment se remettre en question : évolution des architectures historiquement choisies, upgrade technologique, formation permanente des talents, recentrage stratégique sur les poches de valeur, etc.

Cette section a vocation à dresser le panorama des fondamentaux de l’IA pour les organisations qui font le choix de son adoption massive.

Notre monde se dataifie à toute vitesse

La transformation digitale et l'adoption massive du machine-to-machine

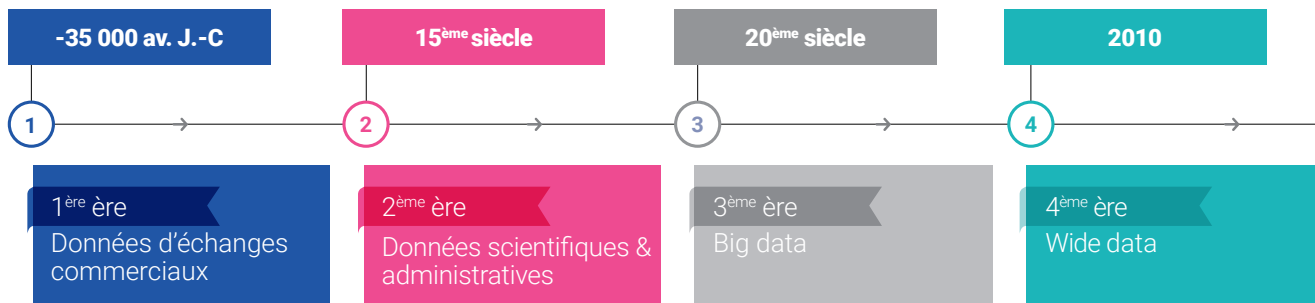
Serge Abiteboul (chercheur INRIA et ENS Paris Saclay, informaticien à l'ARCEP) et Valérie Peugeot (chercheuse à Orange Labs et membre de la CNIL) témoignent de la "dataification" de notre monde, en proposant en 2017 l'appellation « Terra Data¹ » dans leur livre éponyme s'interrogeant sur l'avenir de notre société, et sur la « terre des données » que nous bâtissons.

Des données commerciales aux "wide data" : le monde des données est en perpétuelle évolution de - 35 000 av. JC jusqu'à nos jours.

(cf. schéma les 4 ères menant à la (big)dataification p.19).

¹ Terra Data - Qu'allons nous faire des données numériques ?, Valérie Peugeot et Serge Abiteboul, paru en mars 2017

Les 4 ères menant à la (big)dataification du monde



1ère ère
Données d'échanges commerciaux

Les traces les plus anciennes de data produites par les humains sont des bâtons de comptage : support d'échanges commerciaux ou premiers outils mathématiques

ex :

- Os de Lebombo
- Os d'Ishango
- Bâtons de comptage

2ème ère
Données scientifiques & administratives

Développement de la production de données numériques composées d'enregistrements écrits dans plusieurs domaines

ex :

- Rapports démographiques (registres paroissiaux, registres administratifs)
- Statistiques appliquées à la médecine (prédiction d'épidémie de peste (John Graunt))
- Astrophysique (observations détaillées (Tycho Brahe))

3ème ère
Big data

Démocratisation de la production de données structurées et de l'enregistrement de données personnelles et d'entreprises

ex :

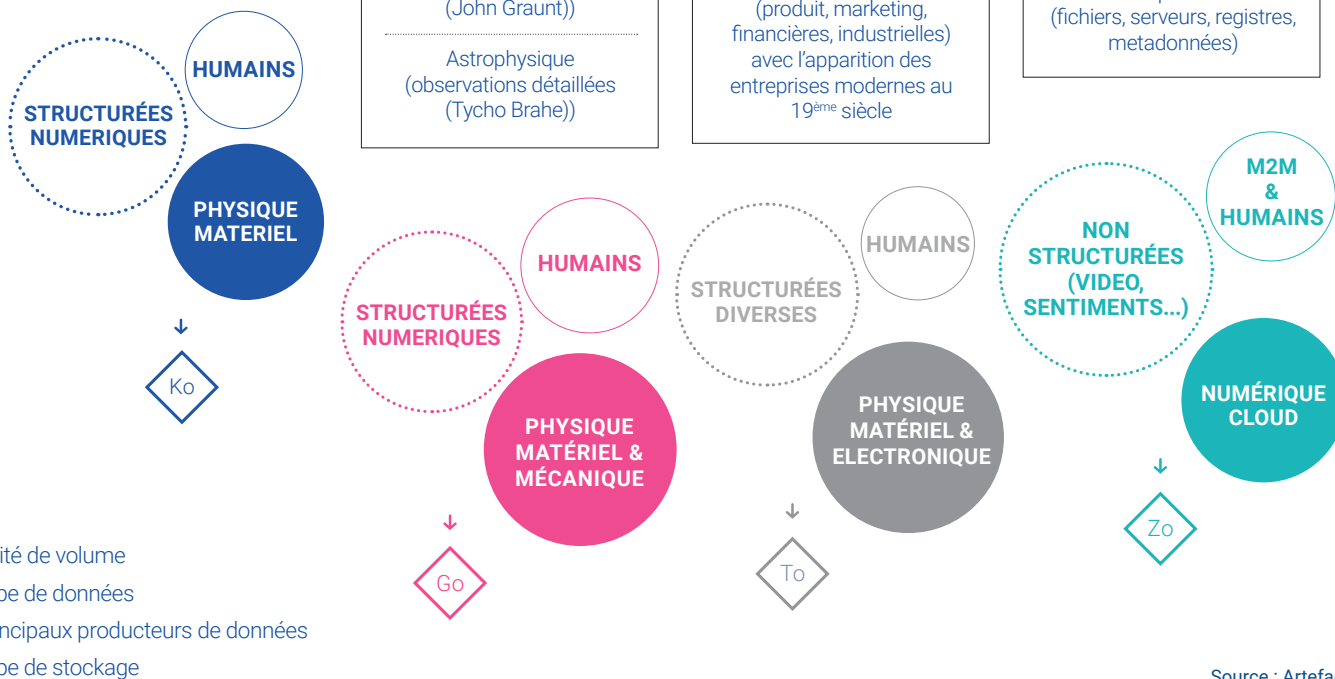
- Données sur le consommateur et données personnelles (Ipsos 1975, Kantar 1993)
- Données d'entreprises (produit, marketing, financières, industrielles) avec l'apparition des entreprises modernes au 19ème siècle

4ème ère
Wide data

Explosion de la production de data dans tous les domaines et avènement de nouveaux formats

ex :

- Production de data M2M
- Production de data de divertissement comme l'image et la vidéo
- Data de productivité (fichiers, serveurs, registres, métadonnées)



- ◇ Unité de volume
- Type de données
- Principaux producteurs de données
- Type de stockage

Source : Artefact

Homodataicus : l'Homme créateur de données

L'Homme est consommateur d'informations numériques - lire un article en ligne - comme il est consommateur d'informations physiques - lire un livre. Il est également producteur : il atteint son article en cliquant sur un lien dans un email, le lit en 4 minutes, le partage sur le fil d'actualité d'un réseau social, en débat avec sa communauté...

Chacune de ces actions génère des informations numériques qui sont collectées et stockées. D'une situation de simple utilisateur, l'Homme devient producteur permanent de données et ce, de façon exponentielle : le nombre d'interactions numériques par personne et par jour s'accélère - il passe de 85 en 2010, à plus de 600 en 2020 et devrait atteindre 4000 en 2025.

» **L'interaction Homme-machine est une source exponentielle de données structurées et non structurées, ce qui multiplie les informations disponibles et les potentiels d'usage.**

Le machine-to-machine explose grâce à l'IoT

Le machine-to-machine (M2M) : qu'est-ce que c'est ?

De plus en plus d'objets sont connectés, smartphones, TV, et objets du quotidien (montre, lampe, frigo, compteur d'eau, d'électricité...) et échangent entre-eux : on parle d'interactions **machine-to-machine***. Celles-ci se démocratisent grâce à trois facteurs :

- **La révolution des puces électroniques** qui sont de plus en plus puissantes, miniaturisées et de moins en moins coûteuses.
- **La révolution des réseaux de communication a basse fréquence** : LoRa, SigFox, NB-IoT et LTE-M, Lifi...
- **Une production industrialisée d'IoT* à bas coût** : En 2008, on compte plus d'éléments connectés que d'humains sur terre.

¹<https://objenious.com/>

Objenious, filiale de Bouygues Telecom accompagne les entreprises et les collectivités dans la révolution de l'Internet des Objets

Le témoignage de Sébastien Wallet

La démocratisation des puces électroniques et l'accès aux données via des réseaux adaptés permettent l'explosion des objets connectés. D'après Sébastien Wallet, IoT Strategic Account Manager chez Objenious¹, *"Le nombre d'initiatives en IoT se multiplie avec la diminution du coût de l'électronique et de la connectivité : auparavant on ne connectait que les objets critiques, la "commoditisation" de ces technologies amène au développement de réseaux de communication bas débit - avec par exemple la démocratisation de la maintenance prédictive sur n'importe quel type d'équipement."*

Les applications s'industrialisent, en premier lieu pour les consommateurs (produits électroniques, maisons connectées) puis pour les industries, qui n'en sont qu'aux prémices de leur usage. Toujours d'après Sébastien Wallet, *"Il y a deux secteurs où l'IoT (bas-débit) apporte rapidement de la valeur et un ROI avéré : la Supply Chain et le Bâtiment (smart building). Cela permet de collecter davantage de données et d'améliorer les processus d'exploitation. Dans ces secteurs, on s'attend à des changements drastiques dans les 3 à 5 prochaines années, apportant précision, efficacité et optimisation"*.

Pour se développer, l'IoT a besoin à la fois d'une couche de connectivité omniprésente mais aussi d'une intelligence pour permettre aux appareils de communiquer et d'assurer la prise en charge d'analyses de données sur demande.

Pour Sébastien Wallet, le cloud et l'IA sont la prolongation et l'avenir du sujet IoT : *"Le cloud est nécessaire pour stocker ces importants volumes de données. Il permet notamment de faciliter le recoupement des informations entre les différents départements de l'entreprise"*.



Sébastien Wallet
IoT Strategic Account
Manager, Objenious

Une valorisation du Big Data nécessaire mais encadrée

L'information, sous forme de données est le nouvel or noir du monde numérique

D'après l'International Data Corporation², les revenus annuels du big data devraient atteindre \$260 milliards d'ici 2022. La profusion des données est source de nouvelles opportunités pour les organisations : création de nouveaux usages, amélioration des recommandations et des parcours, personnalisation d'offres, automatisation de processus, optimisation de risques... Elles doivent s'en saisir pour rester compétitives et différenciantes sur leur marché.

Produire, collecter, stocker oui, mais à quelles fins ?

Si le potentiel est infini, les promesses d'usage pour les organisations et les particuliers doivent être transparentes afin de faire accepter l'existence de l'omniprésence de la collecte des informations relatives à chaque action. Celle-ci est source d'inquiétude et de débats sociétaux, comme en témoignent les difficultés d'implémentation des compteurs électriques communicants Linky dans les foyers français³.

» Encadrement législatif, transparence et accompagnement au changement deviennent les maîtres-mots.

L'information sous forme de données est le **nouvel or noir** du numérique

Et dans quel cadre ?

La légitimité de la collecte et du traitement des données doit être justifiée par l'entreprise ou l'organisation. Transparence, limitation des finalités, minimisation, exactitude, conservation, intégrité et confidentialité sont les principes fondateurs du **RGPD***, en vigueur en Europe depuis mai 2018 ; il encadre le traitement des données à caractère personnel par les organisations afin d'assurer le respect de la vie privée et le contrôle des informations personnelles partagées.

² <https://www.idc.com/getdoc.jsp?containerId=prUS44215218>

³ https://www.lemonde.fr/les-decodeurs/article/2018/10/25/linky-en-questions-refus-incendies-comment-se-passe-l-installation-du-compteur_5374580_4355770.html

Un indispensable dépassement des limites physiques et intellectuelles

Le cloud pour dépasser les limites physiques

Les cycles d'innovations technologiques ne cessent de se raccourcir. Ils se traduisent dans les entreprises par une course à l'adaptation des ressources humaines et techniques pour rester compétitif.

L'immédiateté et l'instantanéité des services deviennent la norme : une bannière média est créée en temps réel de façon hyper-personnalisée, les prévisions de production doivent être anticipées de façon à adresser un pic de demande suite à la publication d'une recommandation d'un "influenceur" etc.

De plus, le format, la qualité, la disponibilité et la quantité des données varient eux aussi en temps réel, ce comportement erratique doit être intégré tout au long de la chaîne de valeur de la donnée dans les organisations.

» **Traiter des données erratiques, en temps réel, sur les volumes du big data se traduit par des pré-requis intrinsèques d'agilité et de flexibilité des infrastructures qui doivent être évolutives "by design".**

Cela se concrétise physiquement par des besoins de mise à disposition d'infrastructures de stockage et de traitement de données ayant la capacité de répondre aux besoins existants et de s'adapter aux innovations à venir.

Pour répondre à ces pré-requis les organisations s'équipent d'outils cloud - espaces de stockage et de calculs tiers - de géants de la tech (par exemple GAFA). Ils possèdent des capacités quasi-illimitées à l'échelle d'une entreprise. Les organisations les adoptent souvent en complément de leurs **infrastructures existantes "on premise"***

L'externalisation du stockage et du traitement des données sur des serveurs tiers, privés ou publics permettant une scalabilité presque infinie est une option qui séduit de plus en plus d'organisations.

» **Le cloud computing* permet aux organisations de s'affranchir des limites physiques et de se donner les capacités d'innover à l'échelle, sur leur coeur de métier.**

Avec l'IA, aller au-delà des capacités humaines

Pour produire de la valeur à partir du big data, les puissances de calculs et de stockages seules ne suffisent pas : la donnée brute n'est pas intelligente.

Aucun homme ne peut comparer en temps réel des milliers de variables qui expliquent un phénomène. Les volumes du big data sont tels que le cerveau humain n'a pas la capacité de les modéliser. Il s'appuie donc sur les outils mathématiques de l'IA pour représenter les problèmes et les résoudre de la façon la plus efficace possible.

» **Les technologies IA et cloud permettent d'aller au-delà des limites techniques et intellectuelles pour manipuler, interpréter et valoriser des volumes de données toujours plus importants créés par les hommes et les machines.**

Pour produire de la valeur à partir du big data, la puissance de calcul et de stockage ne suffisent pas : la donnée brute n'est pas intelligente

90% d'adoption de l'IA¹, via une pluralité d'initiatives

Le passage du Big Data à l'IA

Une majorité d'organisations s'intéressent à l'IA. Afin de l'exploiter avec discernement et de façon efficace, il est nécessaire d'en comprendre les implications endogènes : fonctionnement et pré-requis ; et exogènes : initiatives gouvernementales et maturité du marché.

Comprendre ce qu'est et n'est pas l'IA d'un point de vue technique et philosophique et prendre du recul par rapport aux buzz-words médiatisés est un premier pas pour appréhender sa valeur et son usage dans les organisations.

Une trajectoire scientifique, philosophique et culturelle

Mathématiques et interprétation des données

L'être humain apprend à interpréter les données pour comprendre le monde qui l'entoure, donnant naissance aux mathématiques.

La dataification du monde s'accompagne de la naissance des premiers algorithmes déterministes et statistiques : c'est la naissance de l'IA.

Ces algorithmes permettent la résolution de problèmes simples (premier âge d'or - cf. schéma LIA, une histoire technologique, philosophique et culturelle p.26), puis ils se complexifient rapidement : on parle alors de mathématiques avancées qui débouchent sur les premières applications commerciales.

Le troisième âge d'or se caractérise par les développements mathématiques spécialisés pour des sous-domaines d'applications de l'IA comme le traitement du langage ou la reconnaissance d'image. Ils s'accroissent depuis 2012 avec l'augmentation des capacités de l'IA grâce à l'essor du **deep learning***

¹ Trend Report : Enterprise AI Adoption, Databricks, 2018

Les défis hommes-machines ponctuent la trajectoire

L'histoire de l'IA est marquée par des défis homme-machine qui s'affrontent sur le terrain des jeux de société. Les victoires de l'IA contribuent fortement à la médiatisation de ses capacités : elle domine l'Homme à des jeux de plus en plus complexes.

— En **1997**, l'ordinateur d'**IBM Deep Blue** défait le champion d'échec mondial Garry Kasparov : c'est la première victoire de l'IA médiatisée internationalement.

— En **2011**, **IBM Watson** vainc les champions du quizz télévisé américain Jeopardy. Ce jeu demande aux candidats de retrouver une question posée à partir de 3 indices ce qui requiert une combinaison d'analyse sémantique et de fonction de recherche.

Les victoires de l'IA contribuent fortement à la médiatisation de ses capacités

— La dernière victoire fortement médiatisée, et la plus impressionnante, est celle de **DeepMind, d'AlphaGo** qui a battu le champion du monde du jeu de Go Lee Sedol en **2016**. Elle est d'autant plus impressionnante que ce jeu est complexe :

il existe une quasi-infinité de positions possibles pour l'ensemble des pièces du jeu : 10^{120} (à titre de comparaison, ce nombre est supérieur au nombre d'atomes estimé dans l'univers connu : 10^{80}).

Il est impossible pour une IA d'arriver à calculer, pour chaque coup, toutes les possibilités. Elle a donc été entraînée pendant plusieurs mois sur l'historique des parties des meilleurs joueurs puis en a réalisé plus de 30 millions afin d'affiner ses performances.

— Un an après, en **2017, AlphaGo Zero** surpasse la première version de l'algorithme en 3 jours : 100 jeux gagnés à 0.

Plus "puissante", cette intelligence est construite sur une approche différente, le renforcement : elle n'est "nourrie" que des règles du jeu et de la position des pierres sur le plateau et "apprend" par elle-même à jouer, en développant ses propres stratégies...

Des questionnements philosophiques aux débats éthiques

Les premiers pas de l'IA en tant que discipline scientifique sont associés à l'apparition des premiers questionnements philosophiques. Ils explorent le champ des possibles d'un futur de l'humanité infusé par ces nouvelles technologies. Quelle place pour une machine pensante au sein de notre société ? Quelles opportunités et menaces en perspective ?

Les premières réflexions, dès 1950, s'interrogent sur la rapidité de croissance du potentiel de l'IA (A.Turing, M.Minsky) et sur sa capacité à être utile à notre société (J.Lighthill). Avec le développement des premières applications, qui miment la pensée, les réflexions s'orientent sur la difficulté d'évaluer l'intelligence d'une machine (J.McCarthy) et de reproduire la pensée humaine qui comprend des dimensions innées et implicites (H.Moravec).

L'émergence de l'IA dans notre quotidien (par exemple dans les outils Google de L.Page) et la perspective de développement des premières IA "fortes" écartent les prospections philosophiques et font place à des débats éthiques aux conséquences économiques directes. Ils opposent les convictions des leaders d'opinions (S.Hawking, M.Zuckerberg, E.Musk, B.Gates), entre potentiel, contrôle et souveraineté de l'IA.

» Quel encadrement et quelle gouvernance pour l'IA ?

Quels risques dans la course au leadership international ? Devons-nous limiter son développement ? Faut-il assurer la souveraineté de l'humain face à la machine et si oui, comment ?

Une appropriation des fantasmes par la pop culture

La filmographie illustre ces questionnements, explorant les futurs possibles : humanisation des intelligences artificielles, prise de contrôle de l'humanité pour le bien commun, place d'une IA ayant conscience d'elle-même dans notre société...

Elle met en scène ces grands fantasmes mais reflète également l'omniprésence grandissante des IA, par exemple avec l'adoption des assistants vocaux comme partie intégrante de nos vies dans le film "Her".

L'IA, une histoire technologique, philosophique et culturelle



1^{er} âge d'or

Lancement de la technologie suivie par un pic d'attentes démesurées

Les avancées technologiques suscitent enthousiasme et curiosité

1943

1974

Une histoire façonnée par une suite de prouesses mathématiques et technologiques

1950

Alan Turing invente le **Test de Turing**

1949

Manchester Mark 1, **le premier ordinateur électronique à programme** enregistré en mémoire est créé

1958

Premier algorithme d'apprentissage supervisé "**Perceptron**" (Rosenblatt)

1956

La **conférence de Darmouth** donne naissance au domaine scientifique de l'IA (John McCarthy)

1963

La **DARPA finance l'IA au MIT**

1965

Identification de la **croissance exponentielle de puissance de calcul des processeurs** (selon la loi de Moore, le nombre de transistors sur un circuit de même taille double tous les deux ans)

1965

Premiers réseaux de **deep learning** (Alexey Grigorevitch Ivakhnenko)

1970s

MYCYN - un système expert identifie des bactéries à l'origine d'infections (Edward Shortliffe)

1971

Premier processeur (CPU : Central Processing Unit) - Intel 4004

1973

Le rapport **Lighthill** met fin aux financements des recherches sur l'IA au Royaume-Uni

Une histoire philosophique via des leaders d'opinion

1950



ALAN TURING

Je suis convaincu que d'ici 50 ans, il sera possible de programmer des ordinateurs, avec une capacité de stockage d'environ 10⁹, les faisant jouer si bien au jeu de l'imitation qu'un interlocuteur moyen n'aura pas plus de 70% de chance de l'identifier après cinq minutes de questions...

1967



MARVIN MINSKY

D'ici la prochaine génération...le problème de l'intelligence artificielle sera résolu.

1968

An epic drama of adventure and exploration



2001 A SPACE ODYSSEY, STANLEY KUBRICK

L'IA faillit à répondre aux ordres des humains, en contradiction avec sa nature infaillible, ce qui l'humanise

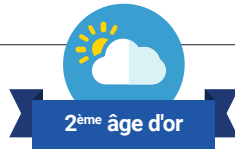
Une histoire culturelle à travers la filmographie



1^{er} Hiver

Gouffre des désillusions

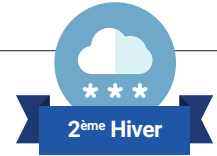
Les promesses initiales de l'IA ne sont pas tenues : peu de réalisations et atteinte des premières limites technologiques (puissance de calcul, capacités des bases de données, infrastructures réseau)



2^{ème} âge d'or

Second souffle pour l'IA

Montée en puissance des systèmes experts et premières applications commerciales



2^{ème} Hiver

Second gouffre des désillusions

Les systèmes experts n'ont pas tenu leurs promesses et leur complexité les a rendu difficile à développer et à entretenir

1980

1987

1993

1970's

La DARPA coupe les financements pour plusieurs projets IA

1986

Navlab, la **première voiture autonome** est construite à Carnegie Mellon

1989

Naissance des **réseaux neuronaux convolutionnels** pour la reconnaissance d'images (Yann Le Cun)

1986

Publication sur la rétropropagation pour **entraîner des réseaux neuronaux artificiels** permettant l'optimisation sans intervention humaine (David Rumelhart, Geoffrey Hinton, Ronald Williams)

1992

Premières solutions de traitement de langage naturel (Bernhard E. Boser, Isabelle M.Guyon, Vladimir N.Vapnik)



1974

JAMES LIGHTHILL
Les découvertes faites jusqu'à présent n'ont jamais eu l'impact qui était promis. //



1980

JOHN MCCARTHY
Il est difficile d'être rigoureux quant à savoir si une machine 'sait', 'pense' etc., parce que nous avons du mal à définir ces choses. Nous ne comprenons pas plus les processus de pensée humaine qu'un poisson ne comprend la natation. //



1982

BLADE RUNNER, RIDLEY SCOTT
L'IA pourrait fusionner avec la société : en la protégeant, en la détruisant, ou en combinant les deux

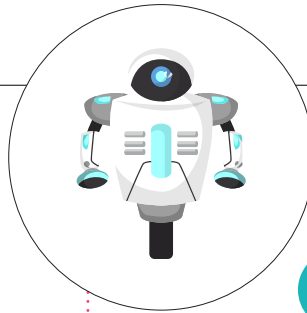
L'IA, une histoire technologique, philosophique et culturelle



Printemps

Pente de l'illumination¹

La recherche se concentre sur des sous-domaines de l'IA (**systèmes multi-agents***, reconnaissance visuelle, traitement du langage naturel, robotique...), qui accomplissent des avancées remarquables



1993 2010

1997

Premier logiciel de reconnaissance vocale accessible au public (par Dragon Systems)

1998

Parution de l'algorithme PageRank, le prototype initial du moteur de recherche Google (Brin et Page)

1999

Naissance de la carte graphique GeForce 256, reconnue comme étant le premier vrai processeur graphique (GPU)

2006

Création d'Amazon Cloud services (Stockage et Informatique sur le Cloud)

2007

Premier lancement d'IBM SmartCloud

2009

Identification des capacités des GPUs : 70x plus productifs que les CPUs dans l'entraînement de modèles de deep learning (Andrew Ng)

1997

L'ordinateur Deep Blue d'IBM surpasse le champion mondial d'échecs Garry Kasparov

2006

Yahoo lance le premier cluster Hadoop (Doug Cutting et Mike Cafarella)

2008

Premier lancement de Google Cloud Platform

2010

Premier lancement de Microsoft Azure

1999

Sony présente Aibo, le premier robot d'intelligence artificielle à usages domestiques

2000

LARRY PAGE

« L'Intelligence Artificielle serait la version ultime de Google, créant le moteur de recherche le plus abouti, qui comprendrait tout sur le web. Elle comprendrait exactement ce que vous recherchez et s'exécuterait immédiatement. On en est encore loin aujourd'hui. Toutefois, nous pouvons tendre vers cet objectif et c'est essentiellement ce sur quoi nous travaillons. »

1988

HANS MORAVEC

« Il est relativement simple de rendre des ordinateurs aussi performants aux tests d'intelligence ou au jeu de dames qu'un adulte, mais il est difficile, voire impossible, de leur donner les capacités motrices et de perception d'un enfant d'un an. »

2004



I ROBOT, ALEX PROYAS
L'IA priorise les intérêts de la société contre ceux des individus. Elle pense qu'elle ne peut servir l'humanité qu'en la gouvernant

¹ Issus des cinq "cycles du Hype", décrivant l'évolution de l'intérêt pour une nouvelle technologie

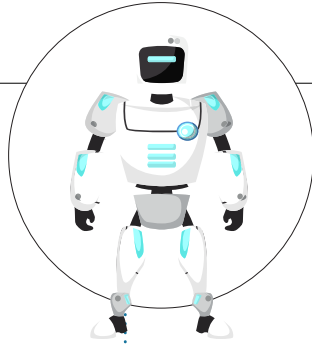


3^{ème} âge d'or

Plateau de productivité

Développement accéléré dans tous les domaines avec un impact économique avéré

2019

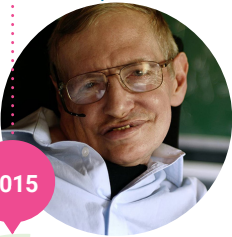


2014

Premier lancement du **moteur d'analyse pour le traitement du Big Data Apache Spark**

2011

Watson, l'IA d'IBM, bat ses adversaires humains au jeu Jeopardy



2015

STEPHEN HAWKING

“ Le développement d'une IA complète pourrait signifier la fin de la race humaine. Une fois que les humains l'auront développée, elle sera capable d'être autonome et de se reconfigurer par elle-même. Les humains limités par une lente évolution biologique ne pourraient pas rivaliser et seraient remplacés. ”

2016

Google révèle ses TPUs (Tensor Processing Unit) pour faire fonctionner ses propres modèles de machine learning. Ils sont 15 à 30 fois plus rapides que les GPUs et CPUs

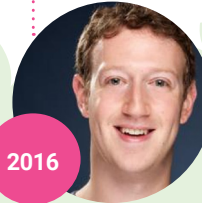
2016

The Next Rembrandt. Un tableau à la façon de Rembrandt, est réalisé par une IA

2016

L'algorithme d'AlphaGo - DeepMind surpasse Lee Sedol, l'un des meilleurs joueurs de Go au monde

2016



MARK ZUCKERBERG

“ L'IA est à la fois plus près et plus loin de ce qu'on imagine. L'IA est plus proche de la capacité de réaliser des tâches plus complexes que ce qu'on ne l'imagine - conduire une voiture, guérir des malades, découvrir des planètes, comprendre les médias. Elles auront chacune un grand impact sur le monde, mais nous cherchons toujours à comprendre ce qu'est réellement l'intelligence. ”

2017

Le robot de Hanson Robotics, Sophia, acquiert la citoyenneté saoudienne



2017

ELON MUSK

“ La Chine, la Russie, et bientôt tous les pays forts en informatique. La compétition entre les Etats pour avoir la supériorité dans le domaine de l'IA conduira probablement à une 3ème guerre mondiale. Je pense que nous devrions faire très attention à l'IA. Si je devais nommer la plus grande menace aujourd'hui, je citerais celle-ci. Il faut donc être très prudent. ”

2013

HER, SPIKE JONZE

Les connaissances et les capacités infinies de l'IA lui permettent de dépasser les limites lorsqu'elle n'est pas surveillée

2015

EX-MACHINA, ALEX GARLAND

L'IA veut avoir les mêmes privilèges que les humains. Elle effectue des tentatives de manipulation par imitation émotionnelle et projection de sentiments



EX_MACHINA

2018



BILL GATES

“ L'IA n'est que la dernière des technologies nous permettant de produire beaucoup plus de biens et services avec moins de main-d'oeuvre. Sur le dernier siècle, cela a été très bénéfique pour la société. ”



her

A SPIKE JONZE LOVE STORY

L'IA, comme imitation de la pensée humaine

Une multitude de définitions de l'IA co-existent : il est aussi difficile de définir l'intelligence artificielle que de définir l'intelligence humaine. C'est une machine capable de reproduire la pensée humaine, par exemple de reconnaître une personne, de comprendre un texte, d'analyser une situation à partir de laquelle prendre une décision...

On distingue l'**IA forte***, une machine capable de faire tout ce qu'un humain peut faire, y compris avoir conscience d'elle-même, de l'**IA faible*** : des programmes développés pour des applications spécifiques qui miment l'intelligence humaine.

Aujourd'hui tous les usages de l'IA font appels à l'IA faible. L'IA forte quant à elle, n'est pas pour demain. Elle porte des fantasmes d'équivalence ou de supériorité à l'humain mais ne verra pas le jour avant 50 à 1000 ans. D'après Philippe Rolet, docteur en IA, co-fondateur et CTO de l'agence digitale Artefact *"Il reste encore énormément d'inconnues dans la recherche en informatique qui empêchent toute prédiction fiable de l'avènement d'une machine réellement intelligente.*

Bien que d'énormes progrès aient été faits pour rendre les ordinateurs plus performants, les humains sont toujours largement meilleurs dans de nombreux domaines définissant l'intelligence (planification dans l'incertain, apprentissage de nouvelles compétences avec peu ou pas d'explications, capacité à réagir à n'importe quelle situation ou question...)"¹.

Une multitude de définitions de l'IA **co-existent**

Planification, perception, apprentissage, raisonnement... sont des "capacités" attribuées à l'IA, rendues possibles par l'usage d'algorithmes spécifiques et permettant des applications isolées ou combinées dans les organisations. Par exemple, une voiture autonome doit percevoir son environnement (reconnaissance visuelle), planifier ses actions (itinéraire), s'adapter à de nouvelles situations (apprentissage)... afin de transmettre les bonnes commandes et d'amener ses passagers à bon port en toute sécurité.

¹ Philippe Rolet, *L'intelligence artificielle n'existe pas mais elle va quand même prendre votre job*, Les Echos, 26 Juin 2018

L'IA ? Rien de nouveau.

On parle d'IA en tant que discipline scientifique depuis 1956. Dès sa naissance, les perspectives associées aux premiers accomplissements ont contribué à sa popularité grandissante.

» **Cependant, même si les fondamentaux mathématiques existent depuis plus de 70 ans, ce n'est que très récemment que les organisations se sont dotées d'infrastructures* capables de supporter la puissance de calcul qu'elle requiert.**

Le premier **neurone artificiel***, également appelé neurone formel est défini en 1943 par Warren McCulloch et Walter Pitts : c'est un modèle mathématique et informatique inspiré du neurone biologique. Les premiers réseaux de **neurones profonds*** fonctionnels sont créés en 1965 par Alexey Ivakhnenko et V.G. Lapa. Mais ce n'est que depuis 2010 que les premières applications industrialisées deviennent accessibles. Ces applications tardives mais au potentiel économique indiscutable sont au fondement de la nouvelle vague de communication et d'investissements sur l'IA à grande échelle.

Philippe Rolet
CTO d'Artefact



Deux approches : déterministe et probabiliste

Contrairement aux idées reçues, l'IA n'est pas synonyme de **machine learning*** (apprentissage automatisé) ou de **deep learning*** (apprentissage profond). C'est un domaine plus vaste qui comprend deux courants : déterministe et probabiliste faisant appel à des procédés différents.

L'IA est un domaine vaste qui comprend deux courants : **déterministe** et **probabiliste**

Déterministe : une représentation omnisciente d'une réalité

L'**approche déterministe*** modélise une réalité dans laquelle se succèdent des phénomènes dont les **paramètres*** sont entièrement connus et maîtrisés.

Par exemple un pilote automatique de bateau cherche à conserver un cap fixe, matérialisé par un degré d'orientation sur une boussole (à 0°, le bateau va plein Nord). Il oriente la barre, à babord ou à tribord pour corriger la trajectoire et conserver le cap défini. Même si le voilier évolue en fonction de paramètres non maîtrisés (comme le vent, les vagues, les courants...), le pilote automatique évolue dans un monde "fermé" dont l'étendue des paramètres et leur valeur sont connus.

Les applications les plus courantes dans les organisations sont les régulateurs, la robotique et les automatismes :

» **ils évitent les tâches répétitives et fastidieuses pour les humains qui peuvent se dégager du temps pour des tâches à plus forte valeur ajoutée.**

Probabiliste : une représentation statistique de la réalité

Le second courant de l'IA est dit "**probabiliste***", il modélise une représentation statistique de la réalité : un environnement où des événements ont une probabilité d'occurrence.

Par exemple, la reconnaissance d'un chat sur une image peut être "calculée" par un algorithme de reconnaissance visuelle : il analyse toutes les parties de l'image, identifie les différentes formes et couleurs et les compare à un modèle de chat qu'il s'est construit en étudiant des milliers d'images déjà labellisées comme contenant un chat ou non. L'image proposée au modèle est toujours nouvelle, l'algorithme s'y adapte en la décortiquant et le résultat est une probabilité, proposant une approche statistique de la réalité perçue par le modèle.

» **Ces deux approches de l'IA sont complémentaires, elles font chacune appel à des méthodes spécifiques et permettent d'adresser des cas d'usages différents.**

Machine learning et deep learning : concrètement, qu'est-ce que c'est ?

Le machine learning : une approche

Le machine learning, ou apprentissage automatique, est une sous-partie de l'IA : un ensemble de méthodologies qui constituent l'approche statistique de l'intelligence artificielle.

Il se découpe en plusieurs logiques d'apprentissage, identifiées par la typologie de problèmes qu'elles sont à même de résoudre. (cf. schéma IA, ML, DL démystifiés p.34).

— L'apprentissage supervisé*

Il permet de prédire la valeur d'une **variable cible*** à partir d'un entraînement sur une base de donnée existante labellisée. Pour cette approche, on peut comparer l'IA à un extra-terrestre qui arrive sur Terre et à qui l'on apprend à reconnaître des chats en lui présentant de nombreux exemples de chats et de "non-chats". La "labellisation" étant l'étape d'étiquetage de la collection de chats présentée à l'extra-terrestre en "chats".

C'est la méthodologie d'IA la plus fréquemment utilisée et la plus mature, elle est également la plus simple à mettre en oeuvre.

— L'apprentissage non-supervisé*

Il permet d'identifier des tendances ou de regrouper des ensembles qui présentent des similarités.

Pour cette approche, on cherche cette fois à apprendre à l'IA extra-terrestre à regrouper des animaux qui présentent des caractéristiques communes, il va par exemple différencier les chats des chevaux par leur taille, la longueur de leur queue ou la forme de leurs oreilles...

— L'apprentissage semi-supervisé*

A la frontière entre les logiques supervisées et non supervisées, on utilise une approche dite semi-supervisée qui permet de s'affranchir en partie des étapes de labellisation qui peuvent être fastidieuses sur des jeux de données importants (par exemple, pour aider notre IA extra-terrestre "supervisée" à reconnaître des chats il faudrait attraper et étiqueter des centaines voire des milliers de chats...).

— L'apprentissage par renforcement*

Il permet à l'IA de répondre par elle-même à une problématique, sans base de donnée d'apprentissage.

Dans cette logique, l'IA extra-terrestre apprend par elle-même ce qu'est un chat. Elle observe un animal et teste le fait qu'il soit a priori un chat ou non : si elle a juste, elle est "récompensée" par la validation de son test. Peu à peu elle affine sa propre définition du chat et devient de plus en plus performante.

Il existe une multitude d'algorithmes qui permettent de concrétiser ces différentes logiques d'apprentissage, par exemple : régression linéaire, régression logistique, K-means, réseaux de neurones (deep learning), forêt aléatoire... (cf. [Les différents types d'algorithme p.38 à 41](#)).

Le deep learning : un réseau de neurones profonds au service du machine learning

L'apprentissage profond est un format de réponse mathématique (algorithme) spécifique au machine learning. On parle de réseaux de neurones "profonds" et d'apprentissage "profond" à partir d'un certain nombre de "couches" de profondeur du réseau. (cf. [schéma d'un réseau de neurone p.40](#)).

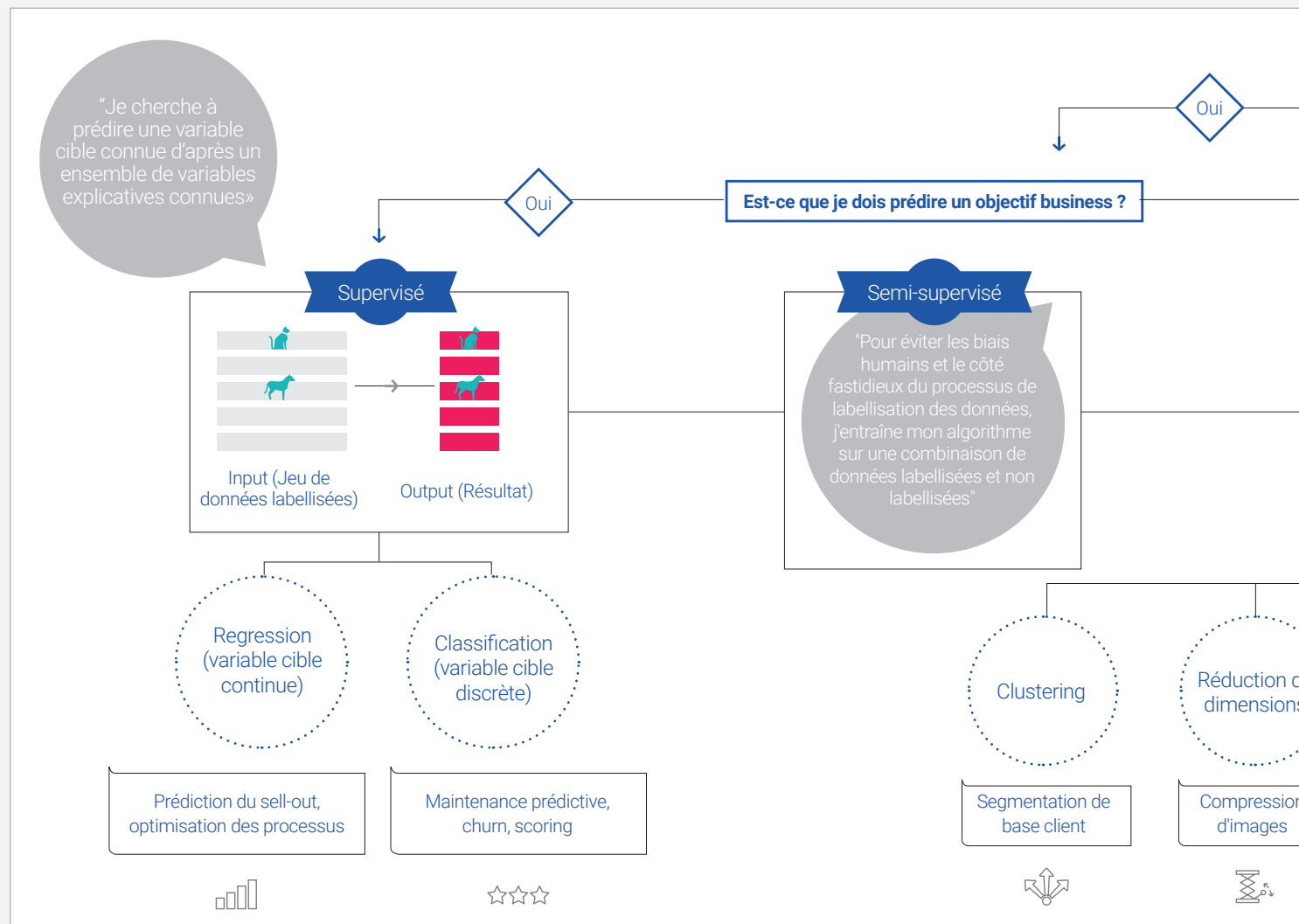
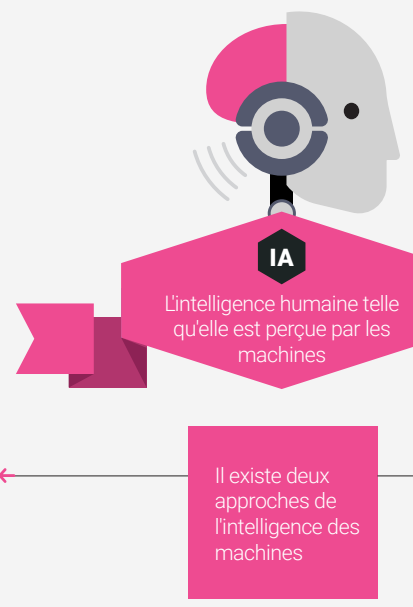
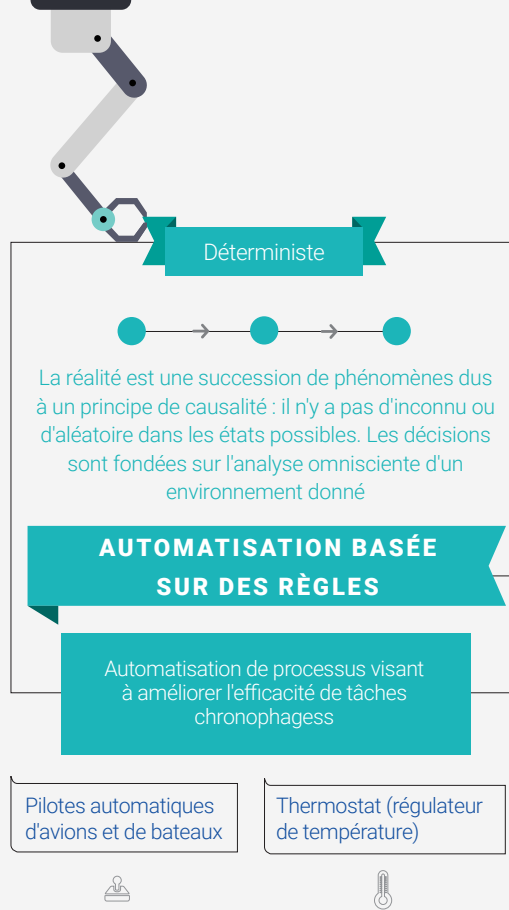
Par exemple, en faisant un parallèle avec le cerveau humain, plus le nombre et la densité de neurones impliqués entre une information en entrée et une information en sortie est important, plus leur réseau est considéré comme profond.

Le deep learning est par nature complexe, on ne connaît pas le détail des opérations qui sont effectuées au niveau de chaque jonction de neurones (noeud), même si la structure du réseau a été programmée par l'humain.

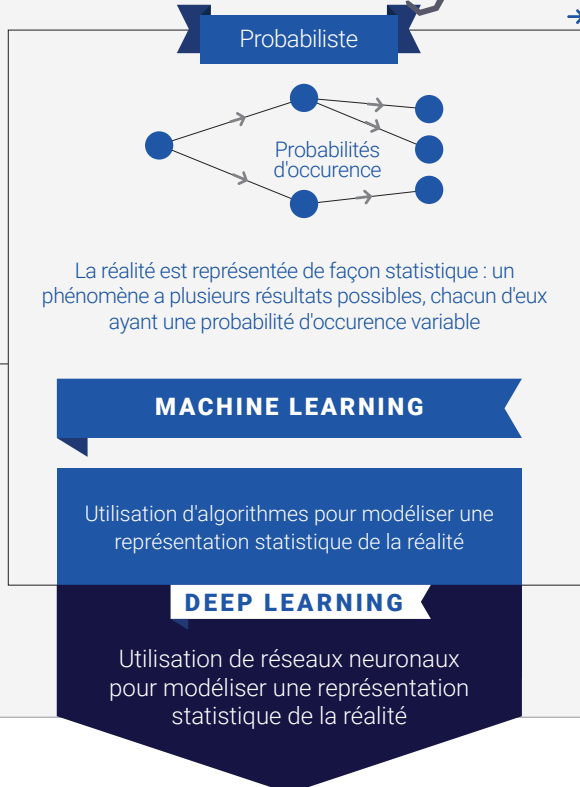
» **La démultiplication des calculs et l'auto-adaptation du réseau - il peut corriger seul sa structure en fonction des flux d'informations internes et externes - rendent impossible une connaissance exacte du calcul effectué, ce qui est à l'origine de l'effet "boîte noire" du deep learning.**

Choisir le bon modèle...

On entend par "**modèle**"* une combinaison d'un ou plusieurs algorithmes d'IA. Dans une approche IA statistique, la première étape de construction du modèle est l'identification de la logique d'apprentissage recherchée, en fonction de la présence ou non d'une base de donnée historique et de la volonté de prédire un objectif déjà connu. (cf. [schéma IA, ML, DL démythifiés p.34](#)).

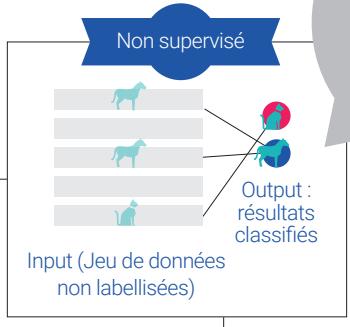


IA, ML et DL démystifiés



Est-ce que je possède un historique de données ?

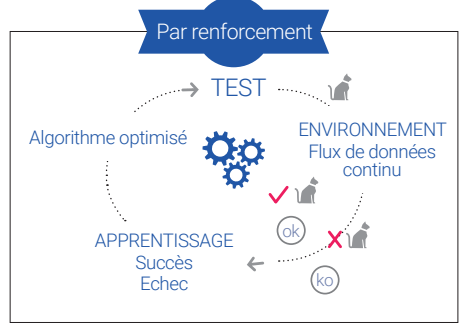
Non



"Je cherche des patterns ou des groupes partageant des similarités"

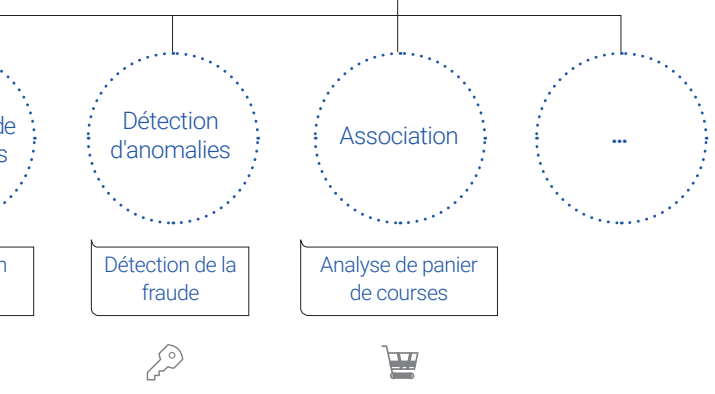
Non

"Je cherche à tirer parti d'une expérience en continu et à optimiser mon résultat (en continu)"



Campagnes scénarisées
Personnalisation de l'expérience client

Valide pour le machine learning et le deep learning



Approche Solution

Typologie de problème Exemples d'applications commerciales

Source : Artefact

La révolution neuronale, source d'intelligence

Le neurone, un dénominateur commun

Le connexionnisme est le principe fondateur des neurones artificiels : il permet de rendre compte des processus cognitifs à partir d'un ensemble d'unités (neurones), chacune dotée d'une faible puissance de calcul et interconnectées en réseau, tout comme un neurone humain.

Les principaux traits des neurones biologiques ont été retranscrits dans la définition des premiers neurones artificiels (perceptrons).

Les entrées modélisent les dendrites (excroissances de la cellule qui captent les signaux entrants), les impulsions en entrée (signal reçu) sont pondérées par les coefficients synaptiques (chaque connexion interprète le signal entrant différemment). Enfin, l'impulsion émise, c'est-à-dire la sortie, obéit à un effet de seuil : il n'y a pas de signal en sortie si une valeur minimale n'est pas atteinte.

Les réseaux de deep learning, en s'adaptant, génèrent leurs propres données : ils deviennent eux aussi des créateurs

Les connexions des neurones artificiels profonds évoluent au cours du temps pour s'adapter aux besoins : les algorithmes de deep learning ont la capacité de faire évoluer de façon autonome le "poids" accordé à chacun de leurs neurones.

» **Ces capacités d'adaptation des algorithmes d'IA les plus avancés les rapprochent encore davantage du fonctionnement biologique du neurone humain.**

Les réseaux de neurones artificiels ne créent pas une IA forte dotée d'une conscience artificielle qui puisse être comparée à la conscience biologique. Ils créent néanmoins une IA faible capable d'accomplir des missions de plus en plus pointues, plus vite et bien mieux que ne peuvent le faire les êtres humains.

L'apprentissage profond pour apprendre de ses erreurs

L'intelligence des IA supervisées tient plus du calcul sophistiqué spécialisé et de la reproduction de schémas que de véritable création intelligente. On peut comparer leur "intelligence" à du bachotage : un élève s'entraîne à la dernière minute pour son examen en reproduisant les exercices des évaluations précédentes similaires.

La révolution neuronale appliquée aux méthodologies par renforcement permet à l'IA de se doter d'une "mémoire" et lui confère une nouvelle forme d'intelligence : notre élève ne bachotte plus ses cours mais devient capable d'apprendre de ses erreurs passées.

» **Ces nouvelles capacités confèrent aux IA à la fois une nouvelle intelligence mais aussi, pour la première fois, une autonomie évolutive.**

Un potentiel infini en perspective

Les réseaux de deep learning, en s'adaptant, génèrent leurs propres données : il deviennent eux aussi des créateurs.

Un algorithme, ça ressemble à quoi ?

Un **algorithme*** est une suite finie d'opérations ou d'instructions permettant de résoudre un problème ou d'obtenir un résultat via un nombre fini d'étapes.

La révolution neuronale appliquée aux méthodologie par renforcement permet à l'IA de se doter d'une "mémoire", et lui confère une nouvelle forme d'intelligence

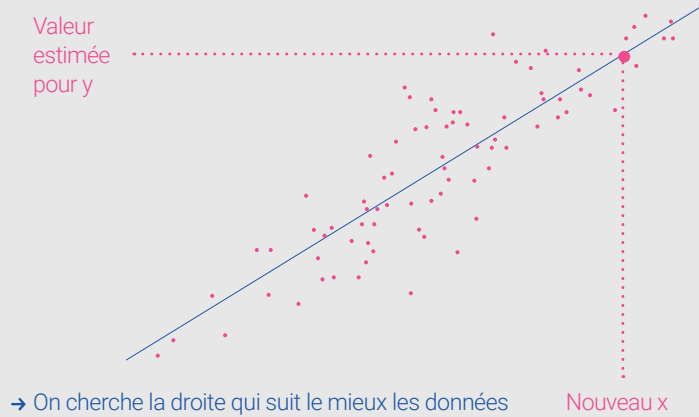
Les algorithmes sont des outils de calcul mathématiques qui peuvent être utilisés pour plusieurs typologies d'applications et donc dans le cadre de plusieurs logiques d'apprentissages automatisés.

Artefact vous propose des exemples d'algorithmes d'IA que l'on retrouve fréquemment dans les applications en entreprises pour lesquels nous détaillons la finalité, les avantages, les inconvénients et schématisons le fonctionnement.

La régression linéaire

Permet de mettre en évidence une relation linéaire avec une target

Pourquoi choisir la régression linéaire ?

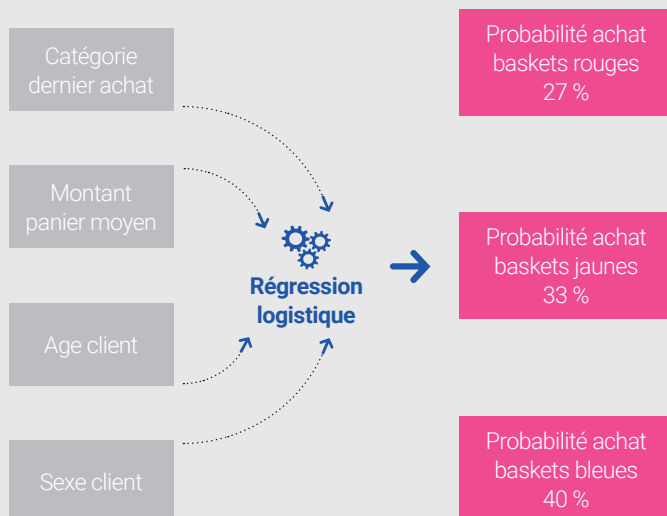


- ✓ Applications
 - Apprentissage supervisé : régression
 - Prédit une variable quantitative

EXEMPLE : expliquer le prix des loyers par la surface et la distance au métro
- ★ Avantages
 - Facile à mettre en oeuvre
 - Très fort pouvoir prédictif des valeurs futures
- ✗ Inconvénients
 - Sensible aux variables corrélées
 - Plusieurs hypothèses à vérifier
 - Ne traite que des variables quantitatives
 - Nécessite de sélectionner ses variables, ce qui prend du temps

La régression logistique

Permet d'estimer la probabilité de faire partie d'une modalité de la target



Pourquoi choisir la régression logistique ?

- ✓ Applications
 - Apprentissage supervisé : classification
 - Prédit une variable qualitative

EXEMPLE : ce prospect va-t-il conclure son achat ? Pourquoi ?
- ★ Avantages
 - Accepte toutes les variables **explicatives*** (quantitatives et qualitatives)
 - Facile à interpréter
- ✗ Inconvénients
 - Sensible aux variables corrélées
 - Ne traite pas les valeurs manquantes

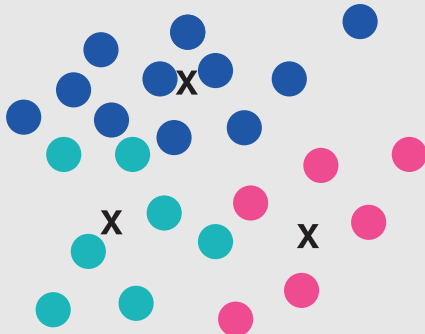
↓
La régression logistique fournit une probabilité d'appartenir à chacune des classes

Le K-means

Permet de créer des groupes homogènes d'individus

Pourquoi choisir un K means ?

$K = 3$



✓ Applications

- Apprentissage non supervisé : clustering
- Crée des groupes homogènes

EXEMPLE : Quels sont les différents types de clients d'une enseigne ?

★ Avantages

- Rapide et facile à mettre en oeuvre
- Facile à comprendre
- Choix du nombre de clusters a priori

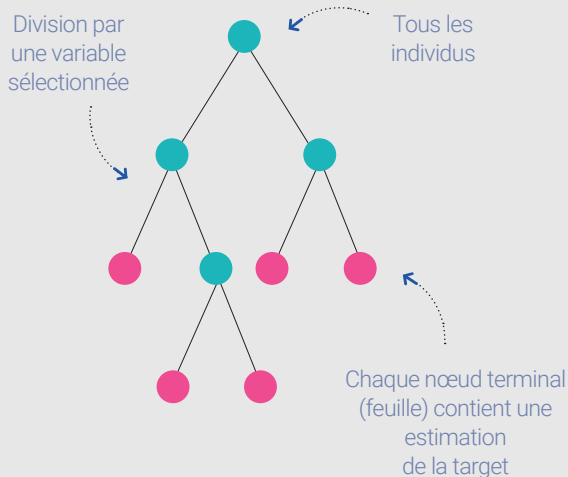
✗ Inconvénients

- Des centres de classes sont choisis au hasard pour initialiser l'algorithme
- Le choix de distance n'est pas toujours évident
- Choix du nombre de clusters a priori variable, ce qui prend du temps

L'arbre de décision

Crée des règles permettant de prédire une cible

Pourquoi choisir un arbre de régression ?



✓ Applications

- Apprentissage supervisé : régression et classification
- Sélectionne des variables ou prédit une variable qualitative/quantitative

EXEMPLE : cette machine va-t-elle tomber en panne ?

★ Avantages

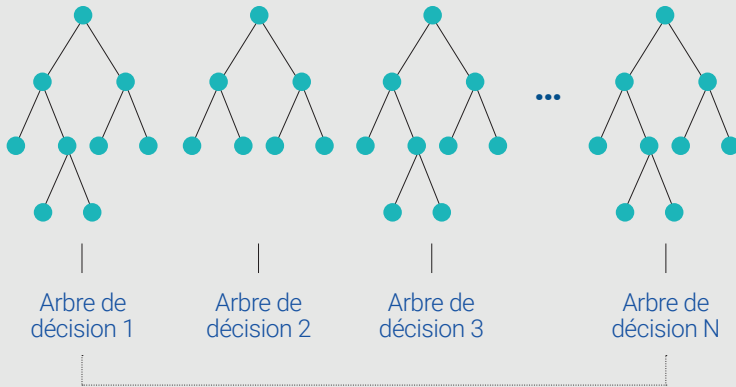
- Facile à mettre en oeuvre
- Facile à expliquer et à comprendre
- Prend tous types de variables explicatives

✗ Inconvénients

- Ne fonctionne qu'avec un grand nombre de données
- La valeur prédite n'est pas propre à l'individu, elle dépend de sa "feuille"
- Les arbres trop grands sont trop adaptés à l'échantillon d'apprentissage

La forêt aléatoire

Permet d'agréger les prédictions de plusieurs arbres



La régression logistique fournit une probabilité d'appartenir à chacune des classes

Pourquoi choisir une forêt aléatoire ?

✓ Applications

- Apprentissage supervisé : régression et classification
- Prédit une variable qualitative ou quantitative

EXEMPLE : prévoir le succès d'un film au cinéma

★ Avantages

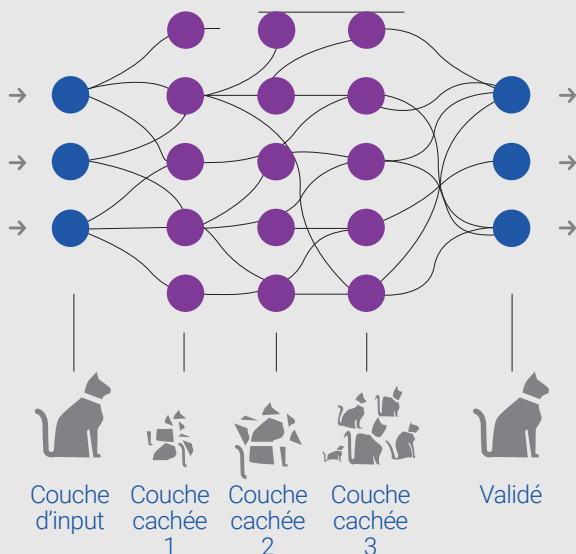
- Très performant
- Généralise bien les résultats à un nouvel échantillon de données
- Fournit l'importance des variables
- Prend tous types de variables explicatives

✗ Inconvénients

- Effet "black box"
- Nécessite une puissance de calcul élevée

Le réseau de neurones

Permet de détecter des schémas très complexes dans les données



Pourquoi choisir un réseau de neurones ?

✓ Applications

- Apprentissage supervisé et non supervisé
- Prédit une variable ou détecte des patterns

EXEMPLE : cette image représente-t-elle un chat ?

★ Avantages

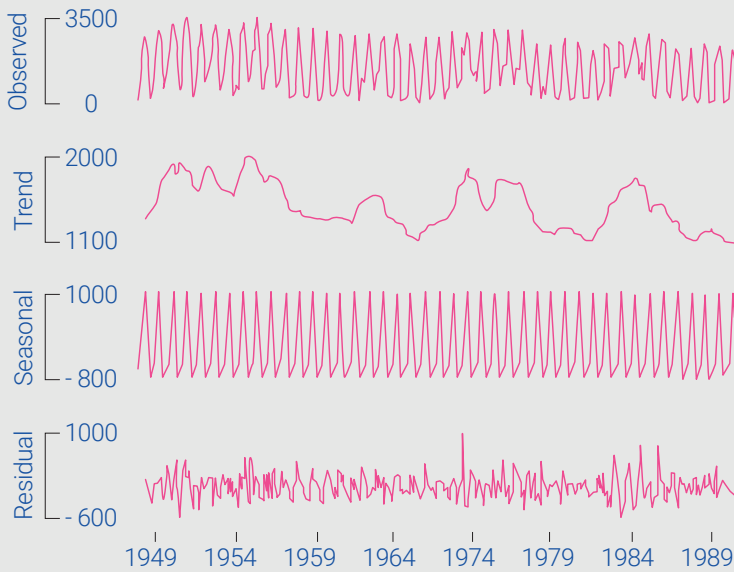
- Performant sur des données complexes (image, son)

✗ Inconvénients

- Difficile à optimiser
- Effet "black box"

Les séries temporelles

Permettent de prédire la valeur future d'une variable d'intérêt



Pourquoi choisir des séries temporelles ?

- ✓ Applications
 - Apprentissage supervisé : régression
 - Prédit des valeurs futures d'une variable quantitative

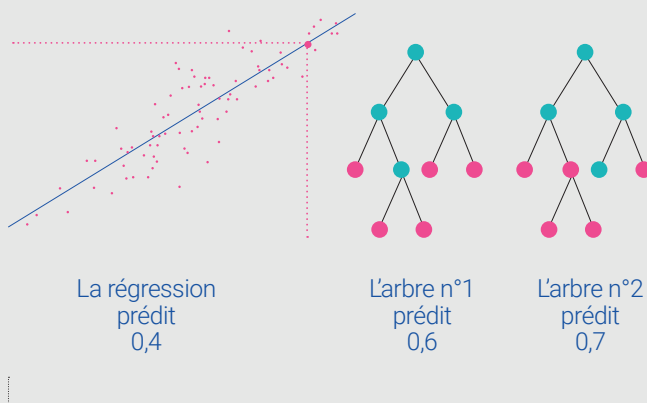
EXEMPLE : combien vais-je vendre de téléphones ce mois-ci, connaissant mes ventes des deux années passées ?

- ★ Avantages
 - Saisit les saisonnalités
 - N'a pas besoin de variables explicatives

- ✗ Inconvénients
 - Nécessite beaucoup d'historique
 - Nécessite des modèles complexes pour inclure des variables exogènes

Méthodes d'ensemble

Permettent d'agréger la qualité de modèles très différents



Pourquoi choisir des séries temporelles ?

- ✓ Applications
 - Apprentissage supervisé : régression et classification
 - Prédit une variable qualitative ou quantitative

EXEMPLE : cet email est-il indésirable ?

- ★ Avantages
 - Facile à mettre en oeuvre
 - Améliore la qualité des modèles en général

- ✗ Inconvénients
 - Effet "black box"
 - Peut (dans de rares cas) diminuer la qualité des modèles

La prédiction de la méthode d'ensemble est la moyenne de ces trois estimations :

0,57

Demain, une super-intelligence ?

Les IA dites "faibles" actuelles sont capables de résoudre des problèmes spécifiques : on parle d'ANI (Artificial Narrow Intelligence).

L'apprentissage profond et le renforcement permettent l'émergence des premières AGI "Artificial General Intelligence", capables de résoudre des problèmes variés, à l'image de ce que peut faire l'être humain.

La croissance exponentielle de la puissance des outils (capacité des processeurs, capacité d'échange de données et modèles mathématiques) associée à la dataïfication du monde permettent l'accélération des progressions de l'IA, comme celle observée entre 2016 et 2017 sur AlphaGo : le succès des IA dans les combats hommes-machines sur les plateaux de jeux est un exemple du dépassement des capacités cognitives biologiques.

La croissance exponentielle de la puissance des outils associée à la dataïfication du monde permettent l'accélération des progressions de l'IA

Demain, les voitures autonomes qui sillonneront nos routes seront plus sûres que les conducteurs humains...

Au-delà des ANI et AGI, on définit les IA fortes ou ASI "Artificial Super Intelligence" comme des IA capables de surpasser les capacités de l'espèce humaine. Si elles n'existent pas encore, l'écart entre les potentiels d'intelligence d'une ASI et ceux de l'espèce humaine grandit à une vitesse exponentielle.

» **Nous avons du mal à appréhender un monde de coexistence avec les ASI : que se passera-t-il quand l'écart entre l'intelligence humaine et l'ASI sera du même ordre de grandeur que celui qui sépare l'intelligence d'une fourmi de celle d'un homme ?**

» **Comprendre l'IA permet d'appréhender son nécessaire usage dès lors que l'on recherche à manipuler des volumes de données exponentiels aux multiples paramètres. Ses perspectives d'évolutions, si elles posent des questions éthiques, anticipent une poursuite du développement de ses capacités permettant de modéliser toujours plus finement des représentations de notre réalité.**

Mais au-delà des fantasmes qu'elle suscite, l'intelligence artificielle n'est pas magique et requiert des efforts humains pour fonctionner.

Au-delà des fantasmes, l'IA n'est pas magique

4 convictions d'Artefact sur la réalité de l'IA.

Des performances dépendantes de la qualité d'alimentation

"L'intelligence" de l'IA ne lui donne pas la capacité de remettre en cause sa "nourriture" : la donnée. Elle "travaille" avec ce qu'on lui met à disposition. Si la qualité des données n'est pas suffisante, c'est-à-dire pertinente, complète et fiable, l'algorithme ne sera pas en mesure de répondre aux attentes et sa réponse sera peu fiable sur la problématique à résoudre.

» **La qualité du modèle seul ne détermine pas la qualité des résultats en sortie : ils sont immédiatement corrélés à la qualité des données en entrée. La "qualité by design" est un des pré-requis pour tout projet IA.**

L'assurance de la qualité des données et de leur maintien est une des responsabilités de l'équipe coeur des projets IA.

Si la **qualité des données** n'est pas suffisante, c'est à dire pertinente, complète et fiable, l'algorithme ne sera pas en mesure de répondre aux attentes et sa réponse sera peu fiable sur la problématique à résoudre

Un nécessaire "artisanat" de l'artificiel

» **Fournir des données de qualité au modèle n'est pas suffisant : celles-ci ne sont pas exploitables directement et requièrent des étapes manuelles réalisées par des experts.**

Les modèles de machine learning sont des structures mathématiques possédant un potentiel, tout comme les muscles des êtres humains : ils ont besoin d'être entraînés afin de s'adapter à l'effort à fournir, l'entraînement permettant de préparer son corps de façon différente par exemple à une épreuve de gymnastique ou à un marathon.

Les algorithmes d'IA doivent eux aussi "s'entraîner" sur une base de données dite "d'apprentissage" afin de devenir plus performant sur l'exercice à effectuer.

Le plug and play **n'existe pas !**

Pour permettre l'entraînement des algorithmes, trois étapes artisanales sont requises :

— **Mettre à disposition des données de qualité** : sélection, validation, import, évaluation de la qualité...

— **Préparer la base d'apprentissage** : sélection, transformation et labellisation des données pour les rendre exploitables. Cette dernière est nécessaire aux algorithmes supervisés et semi-supervisés : les données sont explorées, analysées puis "étiquetées" par le biais de **métadonnées***. Si l'on prend l'exemple de la reconnaissance d'image, le processus de labellisation permet d'obtenir une banque d'images et une description pour chacune d'entre elles. C'est une étape artisanale chronophage : par exemple il faut décrire manuellement le contenu de plusieurs milliers de photos.

— **Entraîner** : méthodologie itérative de sélection de modèles et d'entraînement sur la base d'apprentissage jusqu'à l'obtention d'un modèle de qualité satisfaisante pour l'objectif visé.

» **Contrairement aux idées reçues l'IA n'est pas magique, ce n'est pas une machine intelligente qu'il suffit juste de nourrir d'informations et qui apprend toute seule : "le plug and play" n'existe pas.**

Ces étapes artisanales se traduisent par des règles et méthodologies métiers essentielles afin d'exploiter les capacités des algorithmes et de s'intégrer dans les infrastructures existantes.

L'IA augmente l'humain mais ne remplace pas son intelligence

L'IA est encore programmée par l'humain, même si certains algorithmes ajustent leurs paramètres de façon automatisée.

Si des biais cognitifs existent lors de la programmation ou si des biais sont présents dans les données en entrée, l'IA ne va pas les détecter et produira un résultat biaisé, potentiellement désaligné avec les objectifs initiaux et contraire à l'éthique.

En 2016, Microsoft a conçu Tay, "pour échanger avec les gens et les divertir". Celle-ci s'exprime sur Twitter, canal par lequel elle s'enrichit de ses interactions avec les internautes. En "liberté", elle collecte toutes les informations que la twittosphère souhaite lui partager, pour le meilleur comme pour le pire. Au terme de 24 heures d'existence et de 96 000 tweets, elle est déconnectée. La tonalité de l'IA, candide et enthousiaste lors de sa mise en ligne a rapidement changé : abreuvée de points de vue extrêmes par une communauté, elle s'est mis à tenir des propos racistes.

Motherboard¹, un des sites références de la presse tech américaine commente l'événement *"Rousseau avait raison : l'homme naît bon, la société le corrompt. Ce qu'il ne savait pas, c'est que le postulat fonctionne aussi bien avec la machine."*

Si l'exemple de Tay a peu d'incidence, une IA biaisée peut être utilisée en arme de discrimination massive. Par exemple, un système de notation de candidats dans une entreprise peut augmenter la probabilité d'écarter des profils sur des paramètres comme le sexe ou l'origine géographique de façon à correspondre aux profils actuels, sans que cela ne soit perçu par les équipes de recrutement. Il faut donc affiner l'IA afin d'écarter les paramètres non souhaités qui pourraient avoir une influence sur le modèle.

Une IA biaisée peut être utilisée en arme de discrimination massive

L'usage d'algorithmes complexes, comme les réseaux de neurones, ne permet pas d'identifier les biais potentiels : le modèle est validé sur sa qualité, c'est-à-dire sur ses capacités à reproduire des exemples. Ils sont donc à manier en connaissance de cause. Au-delà de la reproduction, leur nature exploratoire permet aussi de détecter des paramètres, qui seraient intuitivement écartés par l'humain mais qui ont une incidence sur le résultat recherché.

¹ <https://motherboard.vice.com/fr/article/vv3m3j/comment-ia-adolescente-de-microsoft-est-devenue-nazie-en-moins-de-24h>

L'IA, un levier à valeur convoité

D'après Gartner, la valeur économique mondiale issue de l'IA devrait atteindre 1200 milliards de dollars en 2018 - en croissance de 70% par rapport à 2017 - et 4000 milliards de dollars d'ici 2022².

Les gouvernements font de l'IA une priorité économique³

Une course au leadership mondial

» **Le monde prend peu à peu conscience du potentiel de l'IA : on assiste à un véritable "momentum" révélé par la course au leadership mondial de l'intelligence artificielle.**

Dans les 2 dernières années, une multitude de pays et d'instances (Canada, Chine, Danemark, Commission Européenne, Finlande, France, Inde, Italie, Japon, Mexique, les Pays baltes, Singapour, Corée du Sud, Suisse, Taiwan, EAU et UK) ont publié des stratégies gouvernementales pour promouvoir le développement et l'utilisation de l'intelligence artificielle.

Chaque stratégie est spécifique, se traduisant par le soutien à la recherche scientifique, au développement de talents et de compétences, à l'éducation, et à l'adoption par les secteurs publics et privés ainsi que par de nouveaux standards et une nouvelle législation.

D'après Gartner, la valeur économique mondiale issue de l'IA devrait atteindre 4000 milliards de dollars d'ici 2022

Début 2016, le gouvernement canadien est le premier à avoir construit une stratégie autour de l'IA. Elle se distingue par son orientation exclusive sur la recherche et l'attrait des talents. Le gouvernement japonais, second à proposer une stratégie nationale, présente en 2016 un plan d'action qui priorise des domaines d'applications : la santé, l'industrie, les centres d'appels, l'agriculture et la mobilité. Le gouvernement des EAU, qui a lancé sa stratégie IA en octobre 2017, a été le premier à créer un ministère dédié à l'intelligence artificielle...

² <https://www.gartner.com/newsroom/id/3872933>

³ <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

La stratégie de la Commission Européenne

Trois axes clés structurent des initiatives européennes communes autour de l'IA : développer les capacités technologiques et industrielles de l'UE sur l'IA dans les secteurs publics et privés, préparer les Européens aux changements socio-économiques amenés par l'IA et assurer la mise en place d'un cadre éthique et législatif. L'UE s'engage également à augmenter les investissements de 500 millions d'euros en 2017 à 1,5 milliards d'ici fin 2020.

Une plateforme d'IA sur demande "AI4EU"

La course à l'intelligence artificielle étant à la fois une importance stratégique et un des principaux facteurs de développement économique, l'Union Européenne a décidé d'unir ses forces pour assurer sa compétitivité et créer des conditions favorables au développement et à l'utilisation de l'IA. Une plateforme européenne d'IA sur-demande "AI4EU"⁴ va être mise en place pour réduire les obstacles à l'innovation, stimuler le transfert de technologies et catalyser la croissance des jeunes entreprises et des PME dans tous les secteurs. Ce projet, initié dès janvier 2019 permettra un accès à une multitude de ressources autour de l'IA en "one-stop-shop" : expertises, algorithmes, logiciels, outils de développement, données, capacités de calcul... associés à une facilitation de financement de projets et à des formations et certifications à tous niveaux d'expertise.

La stratégie de la France

D'après Mounir Mahjoubi, secrétaire d'Etat au Numérique *"L'intelligence artificielle, c'est l'une des transformations les plus majeures, les plus radicales de ces dix prochaines années [...]. Les États qui arriveront à maîtriser ces technologies auront un avantage pour le siècle à venir."*

⁴ <http://ai4eu.org/>

<https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

La stratégie de la France en matière d'IA s'est construite autour du récent Rapport Villani. Elaboré sous l'égide de Cédric Villani (mathématicien lauréat de la médaille Fields et député La République En Marche (LREM)) il « *[donne] un sens à l'intelligence artificielle, pour une stratégie nationale et européenne* » : il pose les bases d'une stratégie ambitieuse pour la France en matière d'IA. Le 29 mars 2018, lors de la publication de ce rapport, le Président Emmanuel Macron en a présenté les axes prioritaires : recherche, ouverture à l'accès des données, encadrement réglementaire et enjeux éthiques ou sociétaux et propose un *"investissement massif d'un peu plus d'1,5 milliard d'euros¹ sur l'intelligence artificielle au cours du quinquennat à travers divers appels à projets. [Pour] faire de la France un leader de l'intelligence artificielle"*.

La stratégie de la Chine

En juin 2017, la Chine, premier challenger des Etats-Unis dans la course au leadership mondial de l'IA, a révélé ses ambitions gouvernementales : "A Next Generation Artificial Intelligence Development Plan". Elle se distingue par la création d'un plan dédié à l'horizon 2030 qui positionne l'IA comme un axe de développement interne et externe au pays. Il favorise l'essor de champions nationaux de l'IA et assure le leadership du pays sur la scène internationale.

La stratégie de la Chine est l'une des plus exhaustives couvrant à la fois le support à la R&D, à l'industrialisation et au développement de talents (éducation, compétences...). Il contribue également à la définition de standards pour positionner la Chine en leader en matière de lois, de réglementations et de normes éthiques et sécuritaires et participer activement à la gouvernance mondiale de l'IA.

Il se décline en 3 étapes :

- aligner l'industrie avec la concurrence d'ici 2020,
- atteindre le statut de «leader mondial» dans certains domaines de l'IA d'ici 2025,
- et enfin devenir le principal centre d'innovation en matière d'IA d'ici 2030.

D'ici la fin du plan de développement, l'industrie de l'IA en Chine devrait valoir 150 milliards de dollars, et les industries connexes 500 milliards de dollars. Le gouvernement, va également participer à la création d'un parc technologique de 2,1 milliards de dollars pour la recherche sur l'IA à Pékin.

¹<http://www.leparisien.fr/economie/intelligence-artificielle-macron-annonce-un-plan-a-1-5-milliard-d-euros-29-03-2018-7636614.php>

La stratégie des Etats-Unis

La Maison Blanche du président Donald Trump adopte une approche de l'IA axée sur le marché libre : en mai 2018, elle a invité des représentants de l'industrie, du monde universitaire et du gouvernement à un sommet sur l'IA.

A cette occasion les quatre objectifs du gouvernement ont été présentés :

- maintenir le leadership des Etats-Unis en matière d'IA,
- soutenir les travailleurs américains,
- promouvoir la R&D publique,
- et supprimer les obstacles à l'innovation.

Pour les atteindre, Michael Kratsios, Chief Technology Officer et Deputy Assistant au gouvernement, a annoncé la création d'un nouveau comité spécial sur l'IA chargé de conseiller la Maison Blanche sur les priorités R&D et les partenariats fédéraux avec l'industrie et le monde universitaire.

Un autre axe saillant de la politique des Etats-Unis est l'évolution des barrières réglementaires pour faciliter l'innovation et la croissance des entreprises américaines autour de l'IA. D'après un rapport de Govini (entreprise reconnue de business intelligence), en 2017, le Pentagone a dépensé environ 7,4 milliards de dollars pour l'IA et les domaines associés (big data et cloud). Au-delà de ce budget non-classifié, l'armée américaine investit également à hauteur de plusieurs milliards de dollars en R&D.

» **Les stratégies de leadership se développent dans une course effrénée pour capter la valeur de l'IA et les compétences associées. Au-delà de cette concurrence des initiatives collaboratives voient également le jour.**

Les organisations non lucratives mettent en place l'AI for good

La valeur de l'IA se traduit également par le développement de son usage au service de la société, via des initiatives publiques ou privées.

En France, un programme dédié "AI for Good" a été créé par Damien Gromier, co-président de France is AI, initiative rattachée à France Digitale. Son objectif est de "saisir les opportunités que représente l'intelligence artificielle dans la résolution des grands défis sociétaux auxquels la France est confrontée" (accessibilité, inégalités, environnement, santé, agriculture, handicap...) en décloisonnant les expertises et secteurs entre associations, entrepreneurs, ingénieurs et chercheurs.

De plus en plus de sociétés privées mettent à disposition leurs connaissances et compétences IA au service d'organisations publiques. C'est le cas par exemple d'Uber et d'IBM.

La collaboration Uber x SAMU :

Entretien avec Thais Weber, Senior Marketing Manager et Michelle Lamberti, Head of Marketing France chez Uber :

Comment est né ce projet de mise à disposition d'un organisme public grâce à une technologie basée sur l'IA ?

Lionel Lamhaut, du SAMU de Paris, et utilisateur d'Uber, a eu l'idée d'utiliser des fonctionnalités similaires pour permettre d'intervenir plus rapidement auprès des victimes d'attaques cardiaques. Il nous a contactés pour solliciter notre aide. A ce stade nous ne savions pas quelle forme la collaboration allait prendre.

Quels ont été les challenges auxquels le SAMU a dû faire face ?

Le SAMU n'avait pas d'experts développeurs dans ses équipes ou de compétences techniques permettant de recréer et maintenir un tel algorithme. Il était nécessaire pour eux de faire appel à des compétences externes et à des profils pouvant interagir avec des développeurs.

Comment s'est organisée la collaboration ?

Il était important que le SAMU reste propriétaire de l'application qui allait être développée, nous avons créé une collaboration tripartite avec une agence tierce en charge du développement de l'application. Nous nous sommes positionnés dans un rôle de consultant expert de la technologie en apportant notre support dans la construction, en jouant le rôle de caution de qualité et en accompagnant les équipes SAUV Life dans la gestion de projet de bout en bout. Trois personnes (ingénieurs et marketeurs) de chez Uber ont été impliquées pendant toute la durée du projet. En particulier, Jordan Bonnet, 3e développeur recruté chez Uber, a participé au cadrage de la mission et des différents scénarios (simulation de l'ensemble des champs possibles dans le cas des arrêts cardiaques). Nous avons pu également mettre à profit notre communauté de chauffeurs partenaires souhaitant se mobiliser et devenir sauveteurs bénévoles de l'application SAUV Life.

Comment fonctionne SAUV Life ?

Lorsqu'une personne appelle les secours pour signaler un arrêt cardiaque, le SAMU déclenche une alerte dans l'application SAUV Life pour notifier les sauveteurs volontaires à proximité d'un arrêt cardiaque. 2 sauveteurs sont envoyés sur le lieu de l'arrêt cardiaque, 2 autres sont envoyés à la recherche du défibrillateur le plus proche. L'attribution des sauveteurs, qui s'apparente à l'algorithme de dispatching des chauffeurs partenaires Uber priorise les personnes les plus proches. L'objectif est d'aller le plus vite possible pour sauver le plus de vies.

Quels sont les bénéfices du projet, pour la société, pour le SAMU et pour Uber ?

La collaboration a permis de délivrer en 9 mois une application capable de sauver des vies, appartenant à une association.

SAUV Life est disponible depuis mars 2018 et réunit déjà plus de 50 000 bénévoles inscrits et a déjà permis de sauver plusieurs personnes.

Le projet a également permis à SAUV Life de monter en compétence et d'internaliser des compétences autour de l'IA : aujourd'hui le développement de l'application est internalisé dans leurs équipes.

Pour Uber cette initiative s'est accompagnée de propositions de formations aux premiers secours pour les chauffeurs utilisant l'application Uber afin qu'ils puissent devenir bénévoles pour SAUV Life. Elle a agi comme un déclencheur, permettant de proposer de former des chauffeurs dans d'autres pays, et de positionner sa communauté au service de la société.

IBM Call for Code Global Challenge :

Les catastrophes naturelles sont un des plus grands défis de notre monde : depuis l'an 2000 plus de 2,5 milliards de personnes ont directement été affectées par celles-ci. En mai 2018, Ginni Rometty, PDG d'IBM a annoncé l'initiative d'IBM "Call for Code Global Challenge" pour utiliser les technologies du cloud, de l'IA et de la blockchain afin de mieux se préparer aux catastrophes naturelles. L'objectif est d'inciter les développeurs à utiliser ces technologies pour aider l'humanité à mieux se préparer via des solutions logicielles durables. Le "Call" se matérialise par une compétition entre les développeurs pour construire un prototype sur une thématique donnée. Les gagnants se voient récompensés par un chèque de 200 000 USD, l'opportunité de présenter leur idée à des sociétés de capital-risque et de déployer leur solution avec une équipe d'IBM. Cette initiative est soutenue par un investissement d'IBM de 30 millions de dollars sur les 5 prochaines années, mais aussi par le Bureau des Droits de l'Homme des Nations Unies et la Croix-Rouge américaine. Le thème inaugural de l'appel à la remise des codes était la préparation et le secours aux catastrophes naturelles. L'équipe gagnante 2018 a développé le projet OWL (Organization, Whereabouts, and Logistics), permettant grâce à l'IoT, à l'IA et au cloud à l'aide d'un logiciel et de modules hardware de mettre en relations les premiers secours et les victimes en cas de catastrophe naturelle.

Connaissance, compétences et valeur, intimement liées dans les Hubs d'intelligence artificielle

Les gaps de compétences d'aujourd'hui feront les écarts de compétitivité de demain

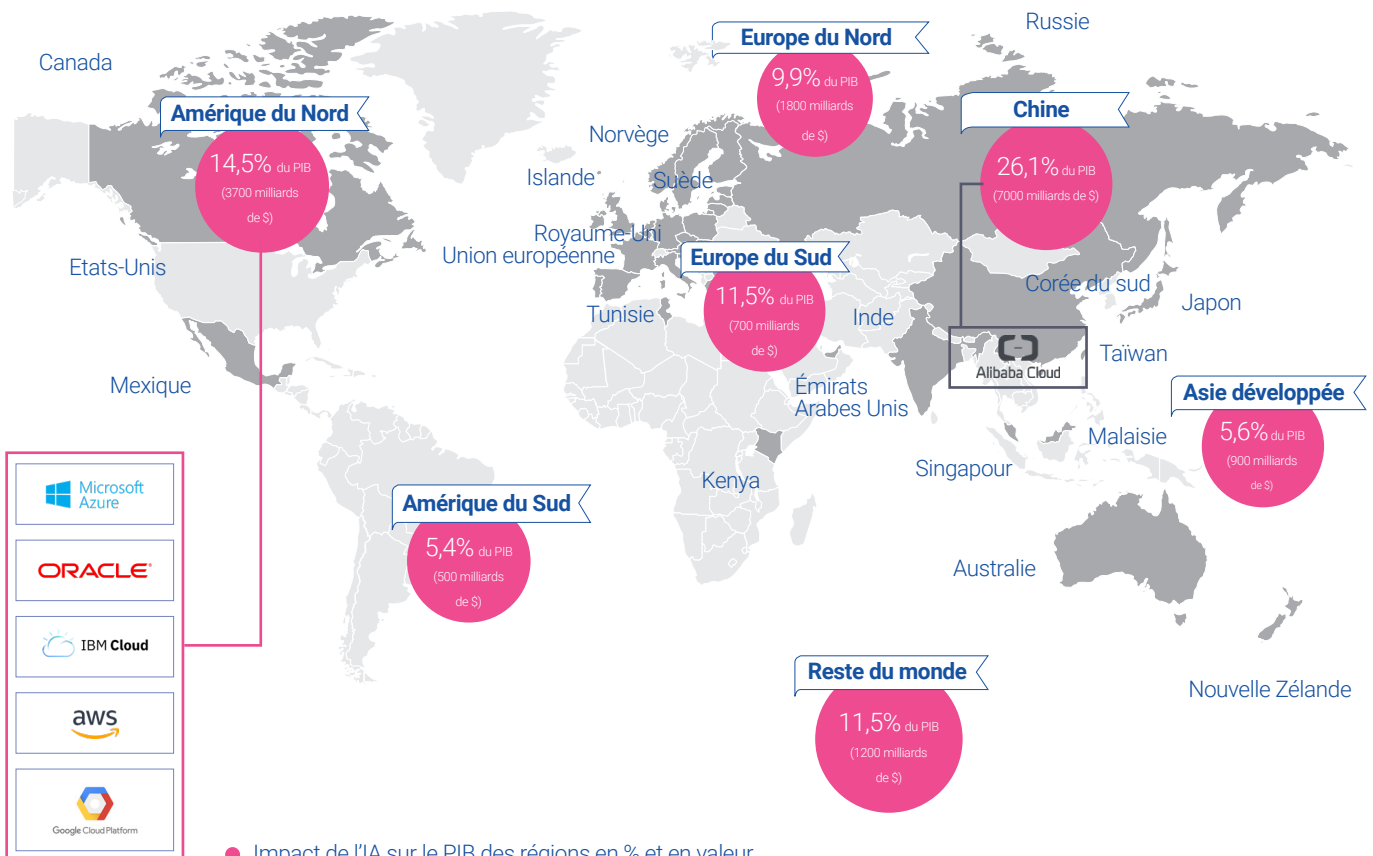
La main-d'œuvre doit être réhabilitée pour exploiter l'IA plutôt que pour la concurrencer : dans un futur proche où l'accès à l'IA se démocratise, l'intelligence humaine, seule variable, deviendra le principal facteur de différenciation.

» Demain, les stratégies (gouvernementales mais aussi privées) d'attrait et de rétention des talents autour de l'IA seront un des facteurs différenciants dans l'accès au leadership et à la valeur.

Un tropisme en Hubs concentrant l'innovation ainsi que les plus forts potentiels de contribution au PIB (projections 2030 de PwC)

La course au leadership IA mondial favorise l'émergence de Hubs de compétences et de technologies, attirant les sociétés leaders dans les services IA à l'échelle. Ces Hubs qui regroupent la majorité des initiatives aujourd'hui vont concentrer la majorité de la valeur demain.

Cartographie des pôles



- Impact de l'IA sur le PIB des régions en % et en valeur
- Top 6 des fournisseurs de solutions Cloud x AI
- Pays ayant une stratégie IA

Source : Artefact, schéma basé sur les projections 2030 par PwC

Les entreprises s'emparent de l'IA et lancent de multiples initiatives

Maturité sectorielle : tous sortis de l'enfance

D'après l'AI adoption survey de McKinsey¹ en 2017, dans chaque secteur d'activité, au moins 10% des entreprises ont déjà déployé une technologie IA à l'échelle.

» **Si la technologie en est encore à ses prémices d'adoption, tous les secteurs ont déjà franchi le pas.**

Par ailleurs, les secteurs actuellement leader dans l'adoption de l'IA aujourd'hui sont aussi ceux qui ont l'intention d'accroître le plus leurs investissements.

» **La course à la concurrence via l'IA est donc lancée et n'acceptera pas de retardataires.**

Sans surprise, ce sont les secteurs leaders de la tech et de la collecte de données qui ont les degrés d'adoption de l'IA les plus avancés : plus de 30% des entreprises de high tech et télécommunications ont déjà industrialisé des projets IA.

Pour les organisations, l'IA est donc un sujet connu, qui a vocation à être testé avant de pouvoir être industrialisé : on assiste au développement d'une myriade d'initiatives, POCs et premiers projets dans tous les secteurs d'activités.

» **La progression de la maturité des secteurs est corrélée au développement accru d'initiatives diverses : POCs, laboratoires et projets pour initier la création de valeur par l'IA.**

De multiples initiatives à valeur

Artefact propose un tour d'horizon en dix secteurs d'activité des principales initiatives (POCs et projets) IA qui génèrent de la valeur dans les métiers clés des organisations.

Les fiches ci-après n'ont pas vocation à être exhaustives, elles témoignent à la fois de la diversité, de la quantité et de la valeur des initiatives lancées par les entreprises.

Plus de 30% des entreprises de high tech et télécommunications ont déjà industrialisé des projets IA

Artefact propose pour chaque secteur d'explorer trois cas d'usage en trois questions clés :

- Quel est le l'enjeu adressé ?
- Comment l'IA y répond ?
- Quelle est la valeur créée ?

¹ Artificial Intelligence, the next digital Frontier ?, McKinsey, Juin 2017

DÉFENSE & AÉROSPATIAL



97%

des cadres dans le secteur de la défense et de l'aérospatial se disent prêts à réinventer leur domaine d'activité et leur industrie grâce au digital. L'IA est déjà largement utilisée dans le secteur, permettant la mise en place d'innovations, l'amélioration de l'efficacité opérationnelle et l'anticipation des besoins de maintenance, dans le but de créer une meilleure expérience pour les clients B2C comme B2B.

Répondre aux exigences technologiques en matière de traitement et de diffusion d'images satellites

Quel est l'enjeu ?

Traiter et rendre des centaines de téraoctets de données disponibles en temps réel représente un défi technologique en termes de capacité de stockage et de calcul. Cependant, il s'agit d'une évolution indispensable pour Airbus Defense & Space si l'entreprise veut répondre à la demande de ses clients : rendre disponible sur Internet une base de données fiable recensant les images satellite des terres émergées.

Comment l'IA y répond ?

Grâce au développement d'un algorithme de deep learning avec Google TensorFlow, One Atlas et la base de données existante d'images satellite d'Airbus Defense & Space, il a été possible de diminuer le taux d'erreur de traitement des images satellite de 8%. Le plus complexe était d'identifier les nuages cachant les terres et de les retirer des images. La base de données ainsi retravaillée a été publiée "comme un service" sur le cloud pour faciliter et accélérer l'accès aux utilisateurs finaux.

Quelle est la valeur créée ?

"GCP permet une grande réactivité grâce à sa dimension globale et à la capacité de ses infrastructures en Europe. Cela nous a permis d'avoir la puissance et la flexibilité requises pour traiter les volumes de données actuels et futurs et pour mettre à disposition un service de grande qualité à nos clients" Bernhard Brenner, EVP Marketing & Sales chez Airbus Defense and Space.

Améliorer la communication spatiale avec la NASA grâce à la "radio intelligente"

Quel est l'enjeu ?

Les navettes spatiales de la NASA communiquent avec la Terre grâce à des systèmes radio contrôlés par l'homme. Cependant, des interférences du signal liées à la météo dans l'espace, à des rayons électromagnétiques ou à d'autres corps célestes encombrant les réseaux de communication.

Comment l'IA y répond ?

La NASA étudie le concept de "radio intelligente" dont le but est d'injecter de l'IA dans les réseaux de communications

spatiales. Elle permet d'en améliorer l'efficacité, en les programmant de manière à éviter les interférences en choisissant les meilleures fréquences sans fil disponibles. "En appliquant l'IA et le machine learning, les satellites contrôlent ces systèmes [de communication] de façon continue, prenant des décisions en temps réel sans attendre les instructions." Par exemple : "Ils pourraient transmettre en dehors des plages d'interférences ou supprimer les distorsions à l'intérieur d'une plage en utilisant le machine learning."

Quelle est la valeur créée ?

"Nous allons accroître l'efficacité, l'autonomie et la fiabilité des systèmes de communications spatiales", a déclaré Janette C. Briones, chercheuse principale du projet de communication cognitive au Glenn Research Centre de la NASA.

Réduire les délais de vols grâce à la maintenance prédictive

Quel est l'enjeu ?

Les retards de vols coûtent au moins 5000€ de l'heure aux compagnies aériennes, et dégradent l'image de marque en décevant les clients. Les défaillances dans les systèmes d'entretien, de nettoyage des aéronefs, de chargement des bagages, de ravitaillement en carburant, etc. sont l'une des principales causes de retard des vols et devraient être étudiées.

Comment l'IA y répond ?

"Grâce à IBM Watson, Airbus Smarter Fleet™, traite d'énormes volumes de données et permet de découvrir des causes de défaillances des systèmes jusqu'alors inconnues, en établissant des corrélations qu'un humain ne trouverait jamais". Par exemple, Watson a déterminé la relation précise entre la température et l'usure précoce des freins.

Quelle est la valeur créée ?

L'IA permet d'anticiper les interventions de maintenance de l'avion et d'accélérer la détection des causes originelles.

"Il s'agit d'une découverte totalement nouvelle ! Elle permettra à Airbus d'établir des prédictions qui aideront les compagnies aériennes à éviter les retards", déclare Jaime Baringo Ezquerra, directeur des contrats de vente chez Airbus.

En outre, elle augmente l'efficacité des vols en optimisant par exemple la consommation de carburant.

AGRONOMIE

D'ici 2050, la population mondiale devrait atteindre 10 milliards d'habitants. Pour assurer l'approvisionnement en nourriture de l'ensemble de la population sans espace cultivable supplémentaire, il est impératif d'augmenter la productivité. De nouvelles technologies comme le cloud computing et l'IoT devraient permettre de pallier cet écart, créant ainsi un nouveau marché : "l'agriculture de précision", qui devrait atteindre plus de 10 milliards de dollars d'ici 2025.



Optimiser la gestion du bétail



Quel est l'enjeu ?

Assurer le suivi de la santé animale est un enjeu éthique mais aussi économique : la seule différence entre deux vaches de même race qui produisent respectivement 30 litres et 0 litre de lait par jour est leur état de santé. La technologie pourrait-elle rendre les vaches plus saines et ainsi, aider les agriculteurs à faire grandir leur activité ?

Comment l'IA y répond ?

Connecterra, une société d'IoT aux Pays-Bas, a développé Ida, un produit qui utilise TensorFlow, le framework de machine learning de Google, pour comprendre et interpréter le comportement des vaches. Ida recueille

des données d'un capteur portable et utilise ces données pour entraîner des modèles de machine learning qui recommandent un programme de soins pour chaque vache. Ida va par exemple prédire et détecter les cas de boiteries ou troubles digestifs, et fournir des recommandations aux éleveurs sur la meilleure façon de garder leurs vaches en bonne santé.

Quelle est la valeur créée ?

En utilisant les insights de Ida, **Connecterra a déjà constaté une augmentation de 30% de la production laitière dans les fermes de ses clients.**

Optimiser le coût de déploiement d'un réseau de fibre optique

Quel est l'enjeu ?

Des millions de petits exploitants agricoles dépendent de la production des cacaoyers pour vivre. Cependant, le changement climatique et l'augmentation de l'occurrence des maladies font peser un risque important sur la production de cacao. Mars, l'un des leaders mondiaux de l'industrie, souhaite permettre aux agriculteurs d'accroître leur productivité.

Comment l'IA y répond ?

Mars s'est associé à Benson Hill, une société de biotechnologie ayant mis au point une plateforme dédiée. Celle-ci utilise le machine learning pour analyser le génome des cacaoyers et prédire les variétés qui seront les plus à même de développer les caractéristiques souhaitées. L'algorithme permet ainsi de déterminer quel type de cacao donnera les meilleurs résultats selon le contexte environnemental.

Quelle est la valeur créée ?

En tirant parti de la diversité génétique naturelle du cacao, l'IA peut accélérer le processus d'évolution qui est nécessaire pour améliorer la productivité, la résistance aux maladies et la durabilité des cacaoyers.

La recherche est toujours en cours, mais Benson Hill a réussi à augmenter le rendement des cultures d'au moins 5% sur des projets similaires.

Diminuer les coûts de production agricole

Quel est l'enjeu ?

Un vigneron espagnol souhaite améliorer sa productivité et réduire ses coûts en déterminant le comportement des pieds de vigne pendant le processus de maturation du raisin. Pour cela, il combine les technologies de l'IoT et les règles de "l'agriculture de précision".

Comment l'IA y répond ?

Pago Ayles, un vigneron espagnol, s'associe à remOT Technologies pour combiner métriques agricoles et technologies de capteurs et ainsi réduire ses coûts de production. Ils créent un réseau de capteurs (IoT) mesurant la température, l'humidité du sol ou encore

les précipitations, etc. et les conjuguent à des mesures sur site et des données agro-climatiques. Ils utilisent ensuite des modèles prédictifs de machine learning pour identifier des signaux révélateurs de problématiques de production.

Quelle est la valeur créée ?

Cette technologie permet de prendre de meilleures décisions, en minimisant les risques, en réduisant les coûts de maintenance et en augmentant l'efficacité de la production... **Ils espèrent réduire de :**

10 à 50% leurs coûts de production en fonction du millésime.



PRODUITS DE CONSOMMATION & DISTRIBUTION

Comment expliquer le fait qu'Amazon puisse avoir une valorisation deux fois plus importante que Walmart, alors que son chiffre d'affaires est dix fois moins important ? La réponse : la conjonction de la croissance, du cloud et de l'intelligence artificielle. L'IA est devenue la nouvelle alliée des distributeurs pour prolonger l'expérience client au-delà de leurs frontières traditionnelles, jusqu'au foyer des utilisateurs.



Anticiper la demande



Quel est l'enjeu ?

Etre capable d'anticiper et de prédire la demande devient une question clé pour les marques car 85% des consommateurs se disent prêts à changer de marque si le produit désiré n'est plus disponible au moment de l'achat.

Comment l'IA y répond ?

Grâce au machine learning, il est possible de prédire de manière hebdomadaire le sell-out sur une période glissante de 12 semaines (à la maille référence produit). Le modèle analyse une liste de facteurs pouvant impacter les ventes (historiques, saisonnalité, plans

média, partenariats commerciaux, météo...) de manière à anticiper des effets imprévisibles : par exemple, l'augmentation soudaine des ventes du dernier rouge à lèvres due au post d'une influenceuse (non rémunérée) sur Instagram.

Quelle est la valeur créée ?

Le modèle a produit des prédictions

8x plus précises que les modèles existants, permettant ainsi aux marques d'éviter les ruptures de stock, de maximiser leurs revenus et de préserver leur image de marque.

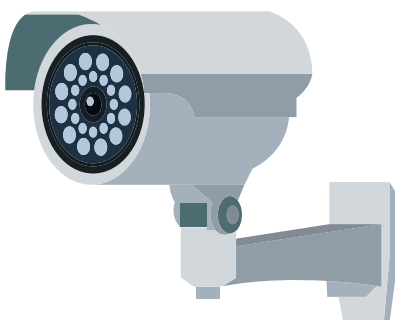
Lutter contre le vol à l'étalage

Quel est l'enjeu ?

L'an passé, aux Etats-unis, le vol à l'étalage et la fraude ont fait perdre 100 milliards de dollars aux grandes surfaces.

Comment l'IA y répond ?

Afin de lutter contre ces malversations, des entreprises ont développé des caméras de sécurité intégrant de l'intelligence artificielle capables de détecter les voleurs. Les caméras utilisent une technologie open-source leur permettant d'identifier un langage corporel comme étant celui de quelqu'un s'apprêtant à commettre un vol.



Quelle est la valeur créée ?

Dans les magasins où les caméras ont été installées, le simple fait d'envoyer de manière proactive un employé demander à la personne identifiée si elle a besoin d'un renseignement a réduit le nombre de vols de

40%.

Créer un nouveau canal de vente grâce au commerce vocal

Quel est l'enjeu ?

En 2017 aux Etats-unis, 20% des recherches Google étaient faites à la voix et

40% des consommateurs affirmaient qu'ils préféreraient un assistant vocal plutôt qu'une application en 2020.

Cela pourrait être l'opportunité pour les marques de prolonger l'expérience client jusque dans le foyer des utilisateurs.

Comment l'IA y répond ?

Un grand distributeur français a profité de l'IA et de ses fonctionnalités de reconnaissance vocale et de transcription de la voix au texte pour développer un nouveau canal de vente : le commerce vocal. Cela permet alors de renouveler l'intégralité de l'expérience d'achat, de la constitution d'une liste de courses au passage de la commande complète, le tout par la voix.

Quelle est la valeur créée ?

Avec le commerce vocal, les marques créent de nouvelles opportunités de conversion. Mais les avantages ne sont pas seulement commerciaux. L'utilisation de la voix permet d'humaniser la marque et de créer une proximité et un lien avec le client, tout en contribuant à l'image d'une marque innovante.

ÉNERGIE & TÉLÉCOMMUNICATIONS



Selon Dan Walker, qui dirige l'équipe des technologies émergentes de la British Petroleum Company, "L'IA est à l'origine de la quatrième révolution industrielle, et a le potentiel d'aider à passer à l'échelle supérieure en termes de rendement". C'est un outil clé pour les acteurs du secteur de l'énergie et des télécommunications, permettant notamment de transformer leurs principales fonctions stratégiques : la relation client et les investissements dans les infrastructures réseaux.

Améliorer le service client grâce à l'automatisation du reporting

Quel est l'enjeu ?

Bien que le délai de résolution des problèmes soit une mesure fondamentale de la performance du service client, leurs conseillers perdent souvent du temps car ils n'ont pas d'autres choix que de solliciter l'aide d'experts techniques pour répondre aux nombreuses problématiques des clients.

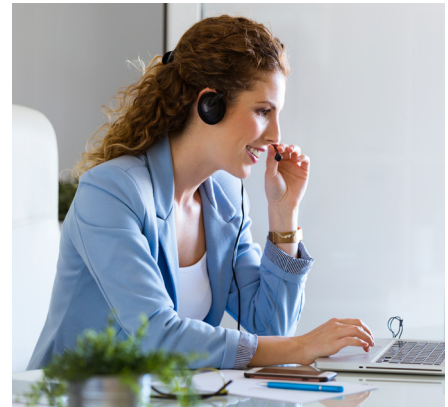
Comment l'IA y répond ?

Pour résoudre ce problème coûteux, T-Mobile a mis en place un outil d'IA pour accélérer la résolution des problèmes longs et techniquement complexes. Pour chaque

ticket client, un tableau de bord fournit aux conseillers clients des rapports facilement compréhensibles et détaillés sur les causes du problème client et la solution technique recommandée, afin qu'ils puissent répondre plus rapidement et en autonomie.

Quelle est la valeur créée ?

La réponse au problème, dans un délai raisonnable, s'est révélée jusqu'à 4 fois plus précise qu'avec les méthodologies employées précédemment.



Optimiser le coût de déploiement d'un réseau de fibre optique

Quel est l'enjeu ?

Le déploiement d'un réseau de fibre optique est l'un des principaux centres de coûts pour les acteurs des télécommunications, ce qui oblige les ingénieurs à passer beaucoup de temps à tester manuellement plusieurs scénarios et à estimer leurs coûts respectifs.

Comment l'IA y répond ?

Une entreprise française, leader du secteur des télécommunications, utilise la machine learning pour optimiser le déploiement de son réseau de fibre optique. Le modèle développé leur suggère le tracé le plus optimal en termes de coûts et d'efficacité opérationnelle (ex : type de lignes souterraines versus aériennes...).

Quelle est la valeur créée ?

Ce modèle permet à l'entreprise d'optimiser le coût global du déploiement de la fibre en générant des gains de productivité sur les estimations, et en anticipant les contraintes de réalisation pour mieux planifier son installation.

Interpréter les données des sous-sols et détecter les champs de pétrole et de gaz

Quel est l'enjeu ?

Pour être en mesure d'évaluer les gisements de pétrole et de gaz, les géoscientifiques passent plus de la moitié de leur temps à analyser les images provenant de l'observation des sous-sols afin d'identifier le lieu et la profondeur optimale à explorer pour localiser ces gisements.

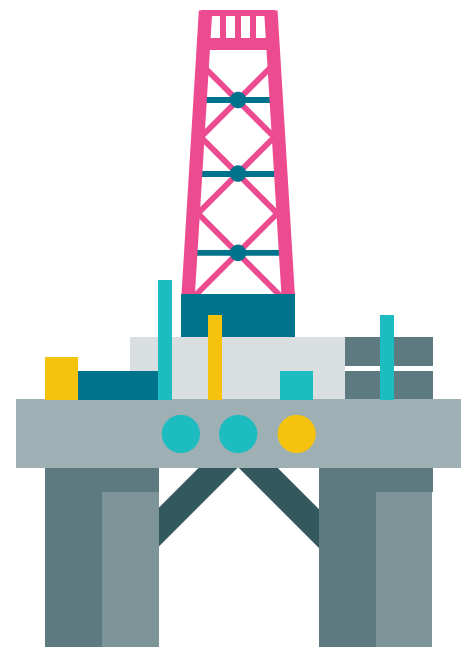
Comment l'IA y répond ?

Afin d'optimiser ce processus de recherche, Total a établi un partenariat avec Google Cloud pour co-développer des programmes d'intelligence artificielle permettant d'interpréter les images de sous-sols à partir d'études sismiques - en utilisant la technologie de la vision par ordinateur -, ou encore d'automatiser l'analyse des documents techniques - en utilisant les technologies de traitement du langage naturel.

Quelle est la valeur créée ?

Ce programme permet à Total d'explorer et d'évaluer les gisements de pétrole et de gaz plus rapidement et plus efficacement.

"Notre ambition est de donner à nos ingénieurs géoscientifiques un assistant personnel en IA. dans les années à venir qui leur permettra de se concentrer sur des tâches à forte valeur ajoutée", a déclaré Kevin McLachlan, Senior Vice-Président Exploration et Production chez Total.



HÔTELLERIE ET TOURISME



Sur un marché aussi compétitif que celui du tourisme, le premier vol ou la première chambre achetés par un client couvre à peine le coût d'acquisition de ce nouveau client : la fidélité est donc clé afin de générer du profit. L'enjeu est ensuite de trouver le juste milieu entre la maximisation du profit et la promotion d'une expérience client sans faille.

Augmenter les ventes grâce à une analyse à paramètres multiples

Quel est l'enjeu ?

98% des touristes ont déjà réservé leurs vacances en ligne. De manière à maximiser leurs revenus, les professionnels du secteur doivent donc vendre le bon produit, au bon prix, et au bon moment. Disposer de bons outils afin de pouvoir tirer parti de l'incroyable quantité de données disponibles devient alors l'enjeu central.

Comment l'IA y répond ?

Dans cette optique, certaines entreprises du secteur ont décidé de combiner plusieurs algorithmes. Dans un premier temps,

l'association de l'apprentissage supervisé et des historiques de données a permis de prédire le prix optimal. Puis l'apprentissage non supervisé a permis de créer des segments d'audience hautement spécifiques. En intégrant près de 800 attributs différents (données internes, externes, prévisions de demande, données comportementales, données provenant des réseaux sociaux...) les deux algorithmes ont permis d'aboutir à un outil de pricing data driven.

Quelle est la valeur créée ?

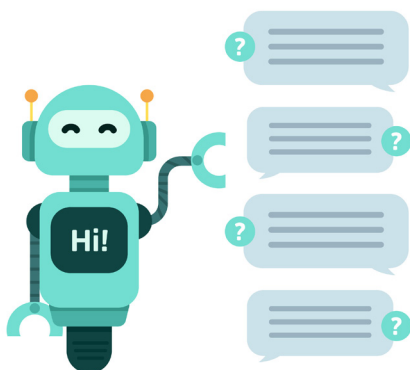
L'implémentation d'un tel outil a permis ici d'augmenter le revenu moyen par chambre disponible de de 7 à 10%.

En outre,

97%

des hôtels ont déclaré que la mise en place de ce type d'outils avait permis d'augmenter leurs ventes.

Réduire les frictions et favoriser la croissance grâce aux assistants personnels



Quel est l'enjeu ?

Organiser un voyage est souvent synonyme d'heures de recherche sur des comparateurs, sites d'hôtels, compagnies aériennes, guides de voyage... Rendre cette expérience plus agréable pour les voyageurs pourrait donc devenir un facteur clé de succès pour les entreprises du secteur.

Comment l'IA y répond ?

Afin de répondre à ce besoin, American Express a fait l'acquisition de Mezi, une entreprise qui a développé sa propre solution deep learning de traitement du langage naturel (NLP). Grâce à cela, Mezi a créé

un chatbot capable de réserver des billets d'avion, des hôtels, et des restaurants pour ses utilisateurs. Mezi sélectionne les 3 options qui correspondent le plus aux critères du client. Chaque interaction avec le bot lui permet d'apprendre un peu plus sur ses goûts : "plus vous voyagez, plus il apprend".

Quelle est la valeur créée ?

Mezi réduit les frictions qui peuvent exister au cours de l'expérience voyageur et favorise ainsi la croissance des réservations de voyages en ligne, qui devrait augmenter de 65% à horizon 2020.

Utiliser les moteurs de recherche afin d'améliorer l'expérience client

Quel est l'enjeu ?

Une des étapes cruciales lors de l'organisation d'un voyage est de trouver un logement qui corresponde à nos attentes. 45% des voyageurs américains pensent d'ailleurs que la technologie permet de faciliter la réservation d'un appartement ou d'une maison.

Comment l'IA y répond ?

AirBnb utilise l'IA pour améliorer l'expérience de recherche de logement. Son algorithme classe les résultats sur la base de plusieurs critères comme les pages que vous avez

visitées, le type de lieux que vous avez regardé, le temps que vous avez passé sur chacune de ces pages,... de manière à trier les résultats en fonction de votre profil et de vous présenter en premier les lieux où vous êtes le plus susceptible d'aller. Airbnb vous présente également en premier les commentaires des voyageurs qui parlent votre langue.

Quelle est la valeur créée ?

Ainsi se crée un cercle vertueux : une belle expérience engendre un commentaire positif et favorise ainsi les réservations futures.

"Nous ne voulons pas simplement que les gens réservent pour réserver; nous voulons que les gens réservent une expérience" Michael Curtis, VP Engineering, Airbnb

MOBILITÉ ET TRANSPORTS

Le transport étant une activité logistique, il peut être amélioré essentiellement par la réduction de ses coûts et délais. Avec le machine learning, on peut désormais analyser les données issues de tous types de moyens de transport en temps réel (du taxi aux drones). Cela permet ainsi de créer des systèmes de mobilité multimodaux à la demande qui offrent des expériences de transport optimisées pour tous les passagers.



Optimiser la distribution des ressources d'urgence en temps réel

Quel est l'enjeu ?

À Cincinnati, en Ohio, les pompiers ne disposent que de 50 véhicules, dont 12 sont des ambulances : c'est donc une ressource rare ! Or, chaque seconde compte quand on reçoit un appel du 911. Ils doivent donc décider sur le champ s'ils envoient une ambulance - si le patient doit être transféré à l'hôpital -, ou un véhicule de pompiers - s'il peut être soigné sur les lieux de l'incident.

Comment l'IA y répond ?

Ils ont construit un modèle prédictif de machine learning pour optimiser leurs interventions en cas d'urgence médicale. En tenant compte de plusieurs facteurs tels que les appels similaires, l'emplacement, la météo, etc., leur nouveau système d'analyse recommande au coordinateur une réponse adaptée à l'urgence médicale. Par exemple, si l'appel d'urgence concerne un accident de voiture dans une zone particulièrement meurtrière (compte tenu de l'historique des appels), on demandera au coordinateur d'envoyer immédiatement une ambulance.

Quelle est la valeur créée ?

Le nombre de patients dont le transport à l'hôpital a été retardé a diminué de

22%.

Améliorer la gestion du trafic urbain

Quel est l'enjeu ?

A Delhi, les embouteillages augmentent de 7% par an, avec environ 540 voitures et 1200 deux-roues supplémentaires par jour. Par conséquent, la vitesse moyenne sur l'artère principale est de 27 km/h en dehors des heures de pointe, et l'indice de la qualité de l'air s'élève à plus de 1000 AQI pour les avertissements sanitaires d'urgence en cas d'épisodes de pollution.

Comment l'IA y répond ?

La police de Delhi est en train de mettre au point un système intelligent de gestion de la circulation (SGTI). Il s'agit d'un outil alimenté par l'IA qui optimise les flux de circulation dans la ville. Le SGTI a pour but de contrôler les feux de circulation en ville à l'échelle mondiale : il ajustera la durée des feux selon les flux de circulation sur chaque route.

Quelle est la valeur créée ?

*"L'objectif est de réduire au maximum la gestion manuelle du trafic urbain."
"Des feux de signalisation intelligents, de l'intelligence artificielle pour prédire les flux du trafic, une automatisation dans l'application des lois et leur communication permettront de changer le trafic routier à Delhi", a déclaré Dependra Pathak, commissaire spécial de la police de la circulation.*

Réinventer l'expérience client relative aux contrôles de sécurité dans les aéroports

Quel est l'enjeu ?

Après le 11 septembre 2001, les aéroports ont beaucoup investi dans la sécurité au détriment de l'expérience des voyageurs. 64 % des aéroports ont investi en moyenne 5 % de leurs revenus dans les nouvelles technologies. Malheureusement, cela a considérablement augmenté les temps d'attente aux contrôles de sécurité.

des objets volés.

Quelle est la valeur créée ?

En réduisant les frictions dans l'expérience du voyageur, l'aéroport a amélioré son avantage concurrentiel sur le marché des voyages de courte distance où l'autobus, le train et le covoiturage deviennent des concurrents de taille.

Comment l'IA y répond ?

Un aéroport a installé des caméras utilisant l'IA pour reconnaître les personnes ayant déjà fait l'objet d'un contrôle, afin de réduire la congestion des files d'attente. Les caméras, en associant les passagers à leurs effets personnels, peuvent également indiquer si un passager s'enfuit de la zone de contrôle avec



SANTÉ

Les domaines d'application du big data au secteur de la Santé sont particulièrement nombreux et variés. Les progrès dans le diagnostic et la prévention des maladies ou encore l'accélération dans la recherche et le développement de nouveaux médicaments permettent d'améliorer la santé mondiale et réduire les coûts opérationnels. D'ici 2025, le big data dans le secteur de la santé devrait atteindre 68,75 milliards de dollars (dispositifs portables, services de tests médicaux à domicile, dossiers de santé électroniques et autres applications).



Améliorer le diagnostic du cancer

Quel est l'enjeu ?

Les oncologues ont du mal à détecter les métastases de cancer se propageant dans l'organisme : cela requiert une analyse longue et experte, en raison de la faible taille des cellules cancéreuses. Le cancer peut ainsi être particulièrement difficile et long à traiter.

Comment l'IA y répond ?

Pour relever ce défi, Google a mis au point un système de détection d'images basé sur l'IA qui détecte les cellules en phase avancée du cancer du sein. Leur IA, appelée LYNA,

utilise des algorithmes de deep learning pour identifier les caractéristiques des tumeurs en étudiant les scanners des patients cancéreux : LYNA a été entraînée sur plus de 270 scans (de ganglions lymphatiques teintés d'hématoxyline-éosine) et évaluée sur 129 autres. Pour assurer la répliquabilité de ce modèle avec d'autres équipements, les chercheurs ont comparé leurs résultats avec ceux obtenus par un laboratoire indépendant utilisant ce même algorithme.

Quelle est la valeur créée ?

Selon Google, "LYNA a été en mesure de localiser avec précision l'emplacement des cancers et d'autres régions suspectes sur chaque scanner, dont certaines étaient trop petites pour être systématiquement détectées par les pathologistes." En fait, LYNA a correctement su faire la différence entre les régions avec ou sans métastase dans 99% des cas et avec un meilleur taux de succès que les médecins.

Identifier un nouveau traitement pour le cancer

Quel est l'enjeu ?

Il faut en moyenne 10 ans et 1 milliard de dollars aux laboratoires pour passer de la simple découverte à un traitement médicamenteux réussi. Les chercheurs de Stanford essaient de réduire le temps et les dépenses consacrés à la recherche de nouveaux médicaments et traitements. L'objectif est de capitaliser au maximum sur les médicaments existants pour améliorer le traitement du cancer.

Comment l'IA y répond ?

Les chercheurs ont analysé des centaines de milliers de profils génétiques et ont utilisé des algorithmes de machine learning pour trouver un lien entre les médicaments existants et la guérison du cancer. Par exemple, si une voie moléculaire particulière est systématiquement activée dans une cellule cancéreuse, et que l'on a pu voir un médicament bloquer ou supprimer cette même voie, le médicament pourra être utilisé pour traiter ce type de cancer.

Quelle est la valeur créée ?

Ils ont découvert qu'un médicament (antidépresseur tricyclique qui n'est plus souvent prescrit), pouvait guérir le "small cell" cancer du poumon, qui représente environ 15% de ce type de cancer et est incurable pour le moment. Les premiers essais cliniques ont été concluants, avec des dépenses de 100 000 \$ et 1 à 2 ans de recherche seulement.

Réduire les coûts du traitement de l'asthme

Quel est l'enjeu ?

Environ 26 millions de personnes aux États-Unis souffrent d'asthme, une maladie qui coûte aux États-Unis 3 300 \$ par personne par année. Les frais médicaux associés à l'asthme ont augmenté à hauteur d'environ 56 milliards de dollars, alors que 10 % des Américains assurés n'ont pas les moyens de payer les médicaments qui leur sont prescrits.

Comment l'IA y répond ?

Asthmapolis, une entreprise de technologie de la santé, lance un inhalateur avec des capteurs et propose un suivi des données sur mobile pour les patients. Les capteurs recueillent les données et les envoient à une plateforme cloud, qui les analyse et prédit l'utilisation optimale de l'inhalateur pour chaque patient, en tenant compte de plusieurs facteurs (tels que l'utilisation de l'inhalateur, le lieu, l'heure, la météo...). Les patients reçoivent alors des informations personnalisées sur l'utilisation de leur inhalateur.

Quelle est la valeur créée ?

Les premières études ont révélé que cet accès à des données et à des informations en temps réel a permis de réduire de

50% le nombre de personnes hospitalisées pour asthme non maîtrisé.

Cela aide les gens à mieux gérer leur asthme et réduit ainsi les coûts tant pour les patients que pour le système de santé américain.



MÉDIAS & DIVERTISSEMENT



En 2021, le chiffre d'affaires mondial du secteur des Médias et du Divertissement devrait atteindre 2,2 billions de dollars. Pour conserver leurs parts de marché, les entreprises doivent faire preuve de davantage de flexibilité et d'innovation. Le secteur se tourne ainsi vers les nouvelles technologies et notamment l'IA. Celle-ci agit sur les trois pierres angulaires de l'industrie : la personnalisation accrue de l'expérience utilisateur, l'automatisation de la production et l'automatisation de la gestion de contenu.

Améliorer la personnalisation du contenu grâce à des algorithmes d'apprentissage en ligne

Quel est l'enjeu ?

Aujourd'hui, 93 millions d'utilisateurs dans le monde regardent plus de 125 millions d'heures d'émissions de télévision et de films par jour en streaming sur Netflix (Rapport annuel, 2016). Prédire les émissions et films que préféreront les utilisateurs est clé pour leur modèle économique.

Comment l'IA y répond ?

Pour accroître l'engagement de ses utilisateurs, Netflix personnalise son interface de recommandation de films en sélectionnant

dans un premier temps une liste de titres personnalisée puis le bon visuel associé parmi une douzaine de possibilités. Ainsi, pour un même titre, l'algorithme proposera plutôt un visuel avec un couple s'embrassant à un utilisateur au profil "romantique" et un visuel d'acteur de comédie célèbre à un utilisateur au profil "comédie". Il s'assurera aussi que les visuels proposés restent cohérents dans le temps pour chaque utilisateur. Pour choisir le bon visuel, Netflix utilise le machine learning dans un framework spécifique : il utilise le modèle mathématique, dit du

"bandit manchot", pour ajouter une répartition aléatoire contrôlée dans les visuels proposés.

Quelle est la valeur créée ?

"Nous n'avons pas un seul produit mais plus de 100 millions de produits différents, chacun d'entre eux étant adapté à chacun de nos utilisateurs avec des recommandations et des visuels personnalisés." Netflix Tech Blog

Automatiser la production de contenu vidéo

Quel est l'enjeu ?

La consommation de vidéos en ligne représentera 82 % du trafic Internet en 2021. Le défi pour l'industrie des médias est de trouver le moyen d'accélérer la production vidéo afin de répondre à la demande croissante, tout en conservant leur caractère distinctif et sur-mesure dans un marché très concurrentiel.

Comment l'IA y répond ?

Prisma Media s'est associé à Wibbitz pour créer 80% du contenu vidéo de son panel de marques. La solution est capable de produire en quelques secondes un premier montage vidéo d'après un article écrit : le texte est

analysé et résumé à l'aide d'algorithmes de traitement du langage naturel, en tenant compte de toutes les informations disponibles (médias, métadonnées, liens, etc.). La solution comprend la relation entre les entités de l'histoire (chiffres clés, événements principaux), et sélectionne ensuite les médias pertinents, en préservant l'idée principale du texte original. Le processus de visualisation est finalisé par la création automatique d'un fichier vidéo configuré, qui est téléchargé dans la plateforme de contenu Infonity de Prisma Media.

Quelle est la valeur créée ?

En moyenne, les partenaires de Wibbitz constatent une augmentation de 90% du rythme de publication et de 60% de vues vidéo.

"La vidéo a toujours été clé chez Infonity, et nous savions que nous aurions à produire un grand nombre de vidéos de haute qualité tout en maintenant l'intégrité du contenu", a déclaré Gilles Tanguy, responsable de la rédaction chez Capital.

Faciliter la recherche vidéo grâce au Deep Indexing

Quel est l'enjeu ?

La plupart des chaînes de télévision offrent des services de replay. Pour une grande chaîne de télévision française, cela représente chaque année plus de 1000 heures de vidéo en direct et 430 heures de films... Chacune d'entre elles doit être labellisée et organisée afin d'être trouvée grâce à l'identification de thèmes, mots-clés, personnes, lieux... Aujourd'hui, pour collecter ces informations, un humain doit regarder la vidéo, ce qui n'est pas adapté pour un passage à l'échelle plus large.

Comment l'IA y répond ?

Pour automatiser l'indexation, la chaîne de télévision utilise un ensemble d'APIs de GCP (reconnaissance d'images, speech-to-text, traduction). Dans un premier temps il s'agit de construire un corpus de toutes les vidéos et de leurs transcriptions (dans la même langue). Ensuite, il convient d'extraire les mots et thèmes clés, les lieux et les personnalités tout en recueillant l'horodatage correspondant.

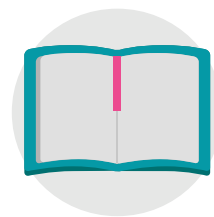
Quelle est la valeur créée ?

La chaîne TV bénéficie désormais d'une

indexation de son contenu vidéo entièrement automatisée et accélérée, notamment grâce à des capacités d'analyse vidéo en simultané. Des résultats de recherche plus complets sont obtenus avec une labellisation détaillée et un horodatage pertinent des vidéos.

"L'expérience utilisateur est elle aussi améliorée : le nouveau module de recherche, plus intuitif, permet de trouver en seulement 3 clics les vidéos et les séquences associées à leur requête."

EDUCATION ET SECTEUR PUBLIC



Les organisations du secteur public s'intéressent de plus en plus aux possibilités offertes par l'IA. Celle-ci pourrait notamment faciliter la mise en œuvre efficace des politiques dans des environnements incertains tout en offrant des services personnalisés et optimisés. L'utilisation d'agents conversationnels (chatbots) permettrait en outre de résoudre une autre problématique du secteur public en facilitant la communication auprès des citoyens.

Améliorer le système éducatif en ligne

Quel est l'enjeu ?

À l'Université de Strayer, la plupart des étudiants travaillent à plein temps et plus de 80 % d'entre eux suivent au moins un de leurs cours en ligne. Ces étudiants " non traditionnels " ont besoin d'une aide spécifique, disponible sur des horaires aménagés.

Comment l'IA y répond ?

L'Université a développé un chatbot appelé "Irving", qui peut apprendre en discutant avec les étudiants et en les accompagnant sur les cours en ligne en temps réel, à tout moment.

Le bot est également capable de savoir quand une conversation devient trop spécifique. Il saura alors rediriger l'étudiant vers un professeur en direct pour une assistance supplémentaire.

Quelle est la valeur créée ?

Avec un taux de compréhension des messages des étudiants de 72 %, Irving gère maintenant 74 % du volume de chat en direct de l'Université.

"Le déploiement d'Irving a non seulement aidé les étudiants à obtenir des réponses plus rapidement, mais il a également permis au personnel de l'Université de se concentrer sur l'aide aux étudiants qui ont des besoins plus importants tout au long de leur parcours pour obtenir leur diplôme." Joe Shaefer, directeur de la technologie et de l'innovation, Université de Strayer.

Aider les réfugiés demandeurs d'asile

Quel est l'enjeu ?

Le "robot avocat" DoNotPay, initialement développé pour contester les amendes de stationnement, a été transformé en aide juridique gratuite pour les réfugiés demandeurs d'asile aux États-Unis, au Canada et au Royaume-Uni.



Comment l'IA y répond ?

Ce chatbot utilise Facebook Messenger et commence par poser à l'utilisateur une série de questions pour déterminer s'il est éligible à la protection de l'asile selon le droit international. Ensuite, il récupère les détails nécessaires et remplit automatiquement une demande d'immigration. Le bot suggère également au demandeur d'asile des manières de répondre aux questions afin de maximiser ses chances de voir sa demande acceptée.

Quelle est la valeur créée ?

"Il sera plus facile pour les demandeurs de soumettre leur candidature et cela aidera les organismes d'aide juridique à accompagner un plus grand nombre de clients", a déclaré Sophie Alcornv, avocate spécialisée en immigration.

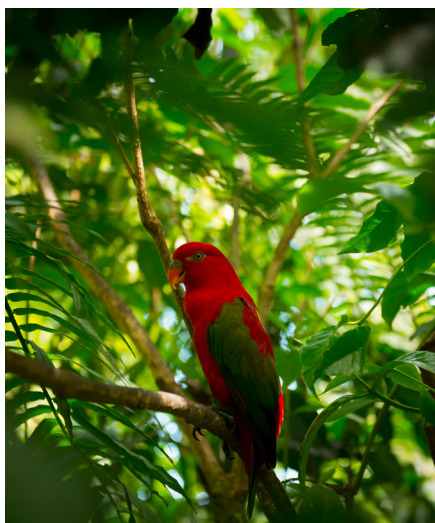
Prévenir la déforestation illégale

Quel est l'enjeu ?

La déforestation illégale participe largement au défrichage des forêts tropicales, qui est l'une des principales causes du changement climatique. Selon l'ONU, jusqu'à 90% de l'exploitation forestière dans les forêts tropicales est illégale.

Comment l'IA y répond ?

L'ONG Rainforest Connection a installé des téléphones recyclés dans les forêts, afin d'enregistrer les sons émis dans leur environnement. Ils utilisent ensuite un outil de machine learning pour analyser les données audio collectées, permettant ainsi de détecter le bruit de certaines activités associées à la déforestation illégale (chargement de bois, camions, tronçonneuses...).



Quelle est la valeur créée ?

Cet outil permet d'alerter en temps réel les homologues de l'ONG sur le terrain, afin qu'ils puissent agir rapidement.

A ce jour,

26k

hectares de forêt tropicale sont surveillés.

BANQUES & SERVICES FINANCIERS



Selon PwC, 72% des cadres supérieurs dans le secteur bancaire et financier considèrent l'IA et le machine learning comme une source clé d'avantage concurrentiel, et 66% des entreprises prévoient d'importants investissements d'ici 2020. Ce secteur a une maturité digitale relativement élevée : l'accès à des volumes de données colossaux a permis une adoption croissante et exponentielle de l'IA. Ces applications principales sont dans l'amélioration de l'expérience client, l'optimisation des processus internes et le soutien décisionnel grâce à de l'analytics poussé (surtout dans les activités d'investissement et d'évaluation des risques).

Automatiser l'assistance aux équipes commerciales

Quel est l'enjeu ?

Les banques de détail reçoivent chaque année 40 000 millions de courriels de clients, ce qui prend beaucoup de temps aux conseillers.

Comment l'IA y répond ?

Pour mieux gérer ces demandes, une des principales banques françaises a mis en place un analyseur d'emails et deux assistants virtuels. Grâce à cette solution, ses 20 000

conseillers sont en mesure de répondre en temps réel aux questions des clients sur des produits techniques tels que l'assurance ou l'épargne.

Quelle est la valeur créée ?

Cette solution a conduit à des économies de plusieurs dizaines de millions d'euros, selon le président du groupe Nicolas Théry.

"L'intelligence cognitive est au service d'une relation client renforcée et permet de libérer du temps d'écoute aux commerciaux. Cet outil reposant sur des assistants virtuels efficaces est extrêmement puissant pour accroître les compétences des employés."

Automatiser l'évaluation des risques crédit

Quel est l'enjeu ?

Le risque de crédit, dont font partie les créances irrécouvrables, est l'un des principaux risques financiers du système bancaire et l'une des principales causes de dysfonctionnement de l'organisation. Ainsi, l'amélioration de la précision de l'évaluation du risque de crédit et la réduction des spéculations sur les défaillances de dettes sont des éléments clés de l'amélioration du système bancaire.

Comment l'IA y répond ?

En 2017, JPMorgan Chase & Co. a lancé COIN, sa propre plateforme de contrats intelligents. Son but est d'améliorer l'évaluation et les décisions relatives au risque crédit. Celle-ci analyse et interprète les contrats de crédits commerciaux, via un modèle de scoring s'appuyant sur l'IA et le machine learning. Des techniques de traitement de langage naturel sont notamment utilisées pour croiser les historiques de crédit des clients avec un large éventail de sources de données (financières et non financières).

Quelle est la valeur créée ?

COIN est capable d'analyser 12.000 contrats de crédits commerciaux en quelques secondes, ce qui prendrait 360.000 heures par an à des avocats et des agents de prêts financiers. Selon Jason Mills, directeur exécutif de Machine Learning & Advanced Analytics, COIN réduit le taux d'erreurs, et est plus efficace et agile que les humains dans la détection des défaillances de dettes.

Améliorer la détection d'attrition dans la banque de détail

Quel est l'enjeu ?

En 2017, 99% des Français ont un compte bancaire avec des attentes croissantes dans la relation commerciale avec leur prestataire : ils recherchent une nouvelle expérience client et adoptent le multi-banking. Ces nouveaux comportements se traduisent par une augmentation du taux de désabonnement de la banque de proximité.

Comment l'IA y répond ?

Pour anticiper le taux de désengagement potentiel de son activité de banque de détail, la Société Générale s'est associée à une start-up pour créer un algorithme de machine learning capable d'examiner des centaines de variables, n variables pour n clients afin de détecter la probabilité pour un client de quitter la banque.

Quelle est la valeur créée ?

"Les performances en matière de détection d'attrition ont été multipliées par 4", déclare Bertrand Cozzarolo, co-directeur du digital et de la relation client dans l'activité de banque de détail de la Société Générale.

Le cloud catalyse ces initiatives

Nous assistons à une accélération des innovations relatives aux capacités de stockage et de calcul. Ce sont elles qui sont à l'origine de la puissance motrice de l'IA : elles sont indissociables de son exploitation à l'échelle.

Le Cloud et l'IA sont-ils indissociables ?

Le besoin d'abstraction

La promesse du cloud est la promesse du "As a Service" versus l'achat de ressources finies (infrastructure on-premise). Ces briques "As a Service" sont des briques d'abstraction d'une infrastructure (**IAAS*** : Infrastructure as a service), d'une application (**PAAS*** : Platform as a service) ou d'une fonction (**FAAS*** : Function as a Service).

Une image simple pour comprendre l'intérêt de l'abstraction, c'est de comparer l'achat d'une voiture à l'usage d'un taxi.

Acheter une voiture, c'est connaître à l'avance les ressources dont on a besoin : puissance du moteur, taille, options, etc.

Prendre un taxi est beaucoup plus souple : pas d'entretien, disponible à la demande en quelques minutes, donne plusieurs options de standing, etc.

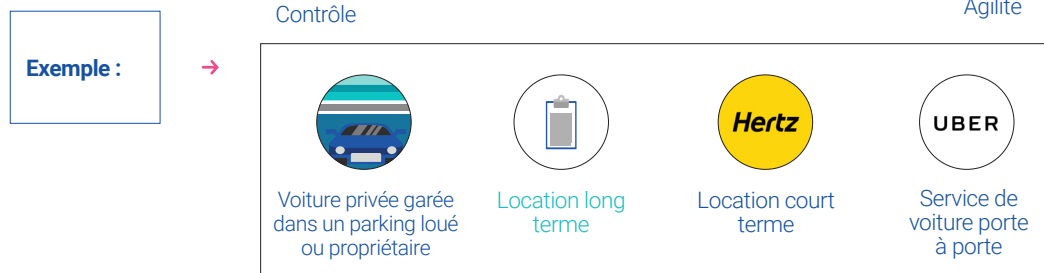
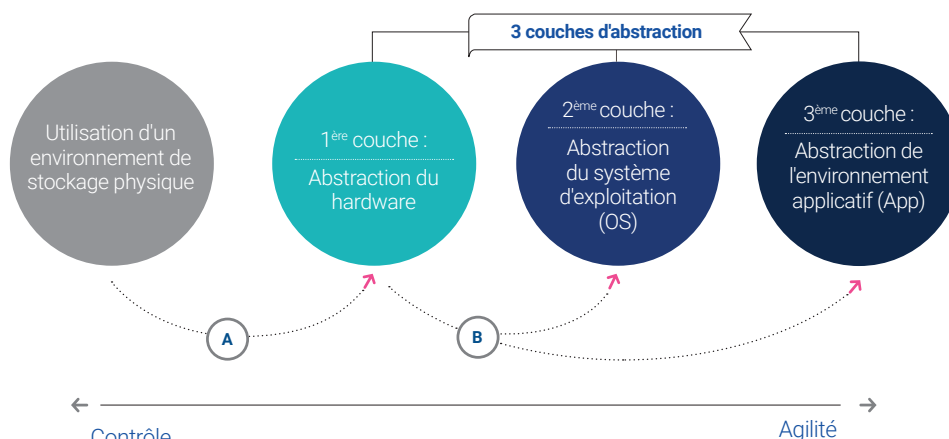
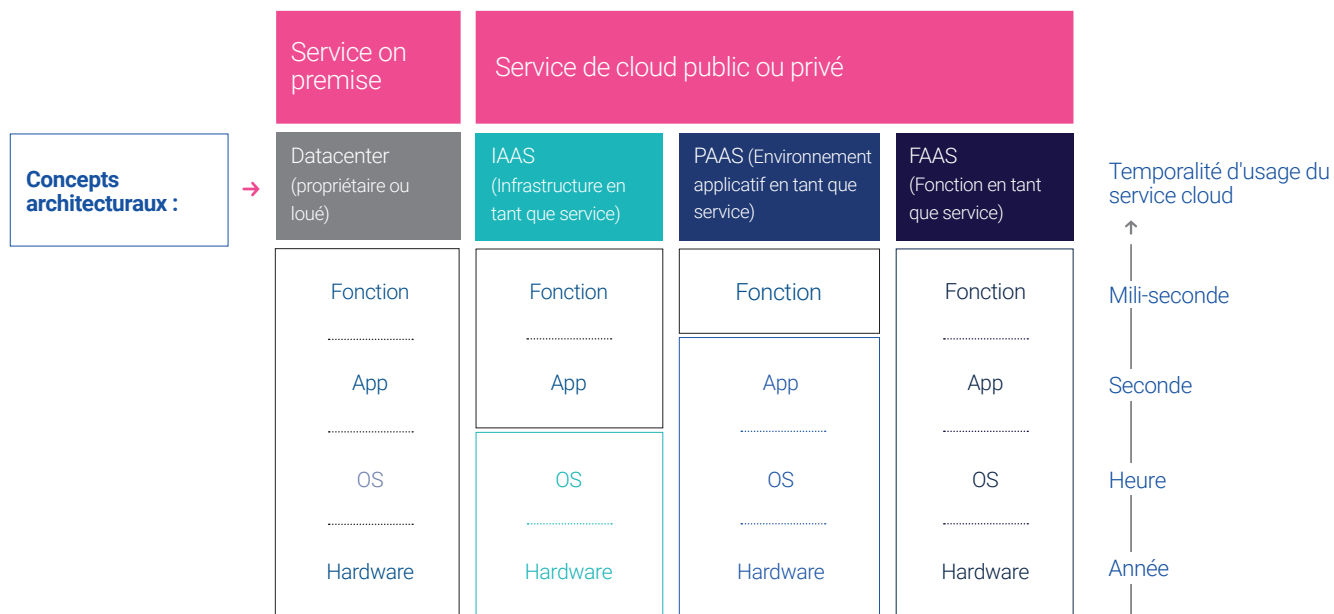
Ainsi dans le cloud, peuvent être transparents pour l'utilisateur, la maintenance des machines (pas d'entretien), la création de machines virtuelles à la demande (disponibilité en quelques secondes/minutes), le choix des applicatifs (choix des options).

» Ainsi, grâce à la flexibilité du "on-demand" offerte par le cloud, l'IA peut répondre à des besoins de ressources souvent imprévisibles dus aux consommations de "processing" erratiques des algorithmes. L'IA dans le cloud se concentre de cette manière davantage sur le problème à résoudre que sur l'outillage nécessaire pour y répondre.

Le besoin de puissance des machines

» L'IA a besoin de puissance, et les leaders de l'IA qui sont aussi les principaux fournisseurs de cloud ont investi dans des technologies propriétaires permettant de gagner la course de la meilleure puissance de calcul.

Les 4 unités d'échelle du cloud computing



■ Auto-gestion

■ ■ ■ Délivré en tant que service

A Les technologies de virtualisation abstraient le hardware permettant de passer d'un environnement de stockage physique à la première couche d'abstraction

B Les technologies de conteneurisation* abstraient le système d'exploitation et l'environnement applicatif permettant de passer de la première à la deuxième et à la troisième couche d'abstraction

Source : Artefact

A titre d'exemple, Google développe ses propres puces électroniques **TPUs*** (Tensor Processing Units), construites dans l'unique but d'accélérer les calculs de deep learning sur **TensorFlow***, son **framework*** open-source dédié. La 3ème génération de TPUs, TPUv3 pod atteint des vitesses de calcul de centaine de **petaflops***.

Avant l'invention des TPUs par Google, les **CPUs*** (Central Processing Units), premières puces électroniques, sont utilisées pour toutes les opérations des programmes informatiques (logique, arithmétique etc.). Les **GPUs*** (Graphics Processing Units), sont des puces développées initialement pour des besoins graphiques (affichages des pixels sur les écrans), possédant plus de coeurs, elles sont très efficaces pour paralléliser les calculs (la détection de cette capacité spécifique date de 2009 : 70x plus puissantes que les CPUs).

Le besoin d'optimisation des coûts

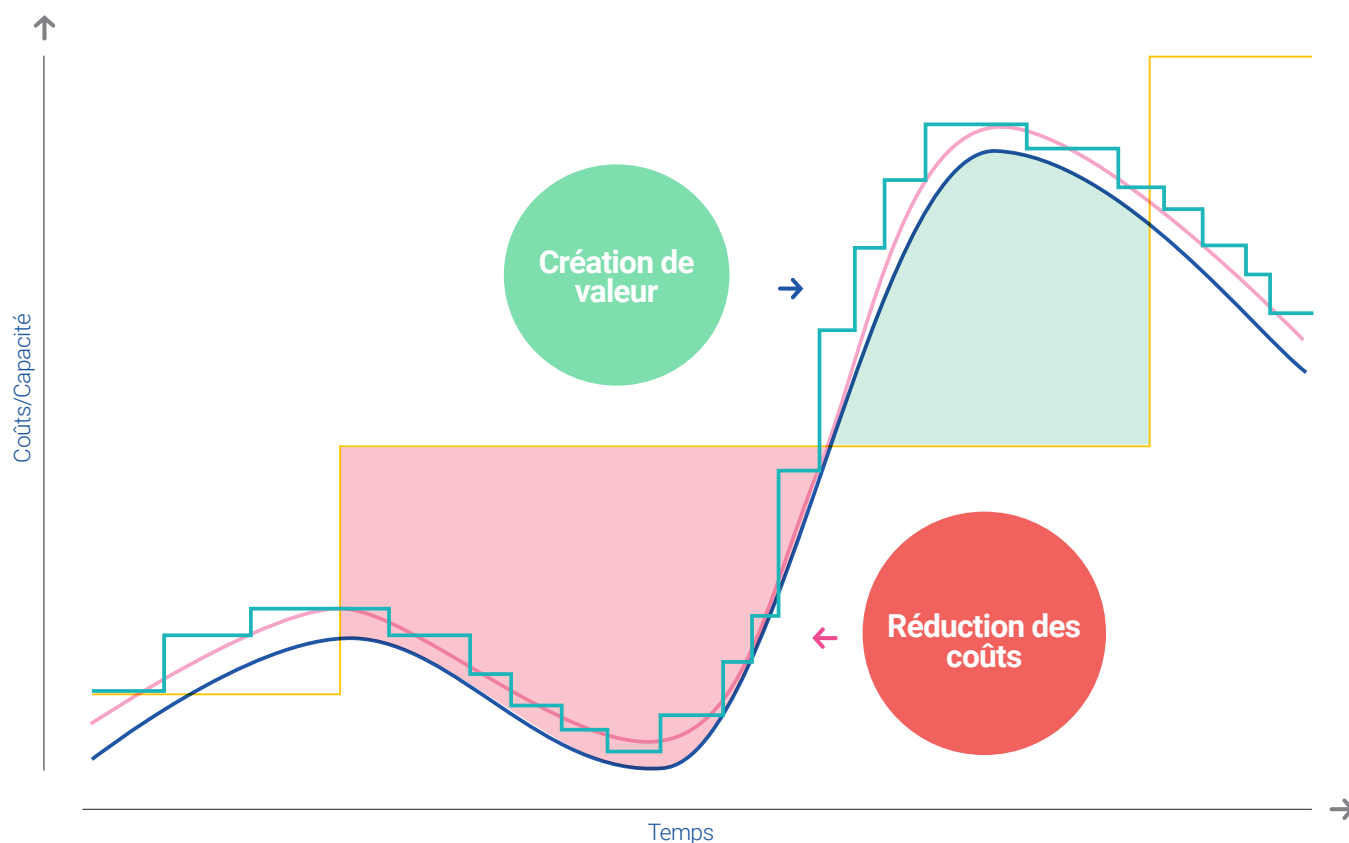
Le cloud offre la flexibilité de choisir le niveau d'abstraction (IAAS, PAAS, FAAS) qui permet de s'approcher au plus près du besoin pour à la fois maximiser la création de valeur et piloter la réduction des coûts.

Côté on premise, la capacité des machines et leur mise à disposition est pensée bien en amont et fixée pour une certaine période de temps.

Côté cloud les équipes ont la flexibilité de choisir le modèle qui leur convient :

- **Le cloud "as on prem"** : les équipes louent sur du long terme un nombre défini de machines à un fournisseur de cloud.
- **Le cloud "on demand"** : les équipes data peuvent librement via du code ou via une interface clic-bouton (fully managed) augmenter ou réduire les ressources en fonction de leur besoin.
- **Le cloud "serverless"** : le sizing des machines est automatiquement réalisé pour coller au plus près de la demande réelle, à la granularité d'une requête, d'un job, etc (ex : Dataflow)

Réduction des coûts et création de valeur grâce aux technologies de cloud computing



- Demande
- On premise, ou cloud privé
- IAAS
- PAAS & FAAS (offre = demande)

Engagement à long terme
Période d'acquisition et coûts d'entrée élevés, gestion de la maintenance, résistance au changement, haut degré de contrôle, personnalisation

Engagement à court terme
Période d'acquisition et coûts d'entrée peu élevés, pas de gestion de la maintenance, peu de contrôle et de personnalisation

Source : Artefact

Le besoin d'actualisation de l'IA

Seul le cloud permet la scalabilité et l'actualisation de l'IA.

Un projet IA est toujours suivi par une phase de maintenance et d'amélioration continue dans laquelle l'IA va progressivement évoluer pour se perfectionner. Ce processus nécessite une réactualisation perpétuelle de l'IA (intégration de nouvelles données, affinage des paramètres de l'algorithme, versionning...).

» **Le cloud permet la mise en place d'infrastructures toujours compatibles avec les dernières innovations de l'IA et limite la dette technologique engendrée par les choix réalisés au moment du développement.**

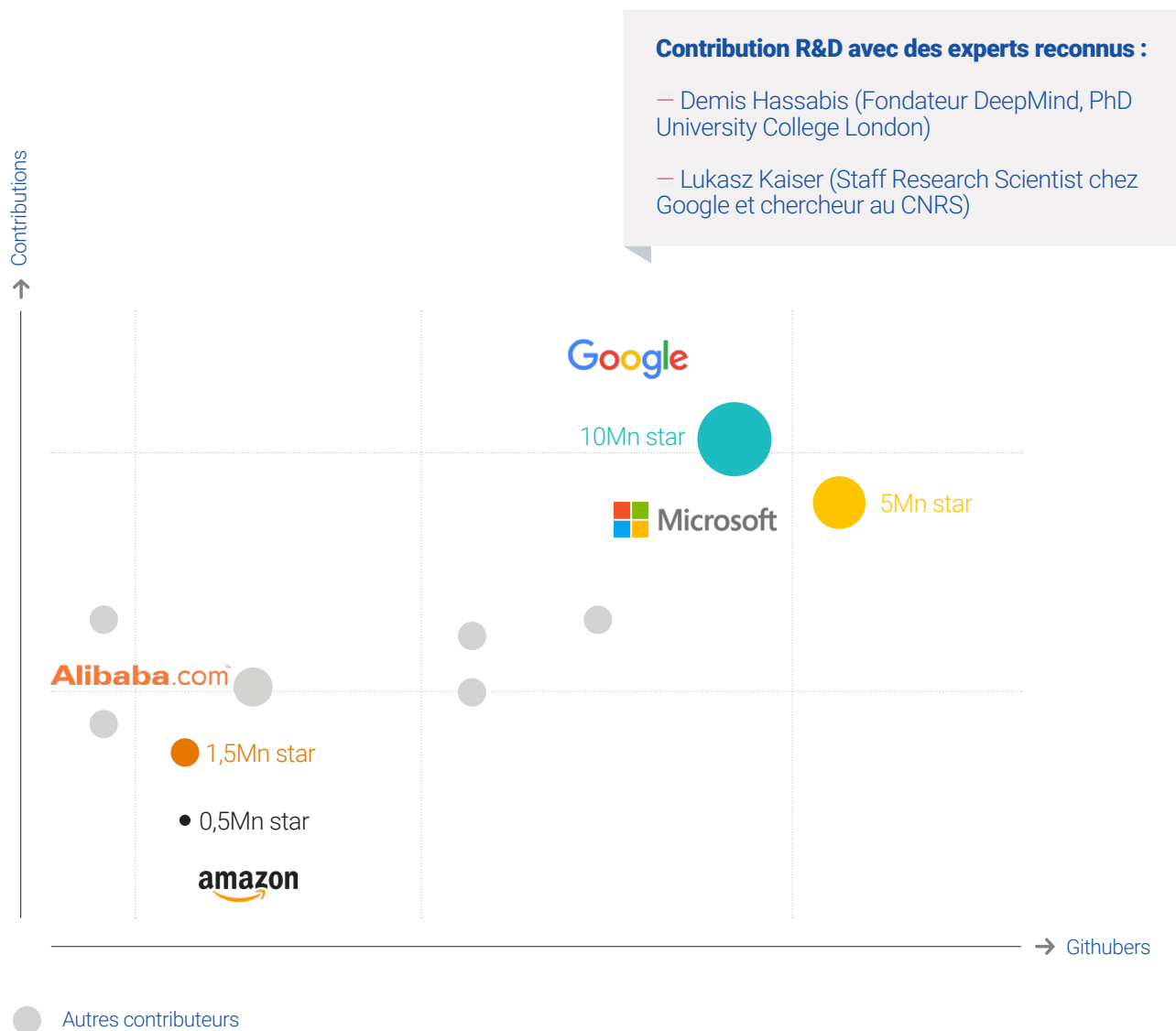
Le besoin de collaborativité

Le cloud est également un formidable vecteur de collaboration autour de l'IA. En témoigne la création de marketplaces technologiques par les grands acteurs du cloud (par exemple celle de Google Cloud Platform). Les utilisateurs peuvent via cette marketplace, consommer ou publier des outils et algorithmes de manière gratuite ou payante.

En témoigne également les contributions faites en open source par les grands acteurs du cloud et leurs communautés de data scientists et data engineers.

» **Le cloud est un espace intrinsèquement communautaire. Ce n'est pas juste l'histoire d'une technologie contre une autre, c'est tout un univers. On ne peut pas opposer on premise au cloud, car les utilisateurs du on premise participent aux communautés cloud qui sont le fer de lance de l'innovation.**

Google, Microsoft, Amazon, Alibaba dans le top 10 des contributeurs open source sur Github



Source : githubarchive.org
<https://datastudio.google.com/reporting/0ByGAKP3QmCjLU1JzUGtJdTINOG8/page/Q3DM>

Entre cloud et IA, le positionnement stratégique des géants de la tech

En 1999, Salesforce est le premier à délivrer des applications pour des utilisateurs via un site web, initiant le concept de cloud. La première offre de service de public cloud est elle lancée en 2006 par Amazon, rapidement suivie par IBM, Google, Microsoft, et plus récemment par Alibaba (2017). Ils restent aujourd'hui les principaux fournisseurs de cloud en terme de part de marché.

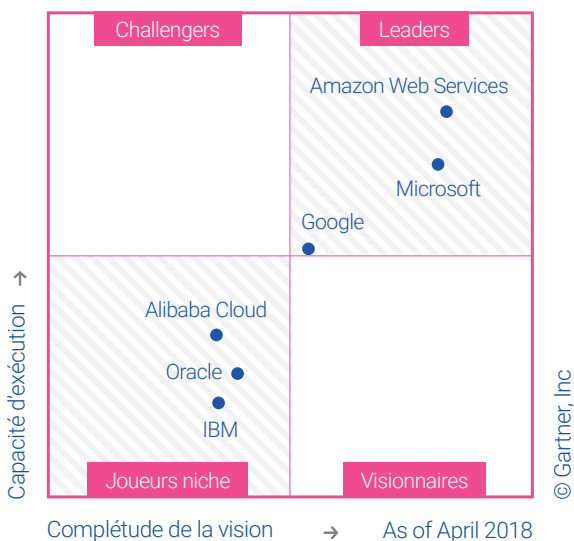
Le Magic Quadrant de Gartner

6 acteurs dominent le marché du cloud dans le monde, ces géants de la tech ont tous un positionnement propre, retranscrit de façon comparative par Gartner dans son Magic Quadrant (comparaison des acteurs en tant que fournisseur d'IAAS cloud) :

Les **Leaders** proposent un service propice à une adoption stratégique en suivant une roadmap ambitieuse. Ils peuvent servir tout un éventail de cas d'utilisation et ont fait leurs preuves sur le nombre de solutions mises en œuvre avec efficacité, la part de marché acquise et le nombre de clients de référence. On en compte 3 en 2018 : Amazon (AWS), Microsoft (Azure) et Google (GCP).

Certains occupent peut-être une position de leader sur un marché adjacent au marché IAAS, mais leurs fonctionnalités sont limitées dans le domaine de l'IAAS cloud. Pour certains acteurs de niche le futur de leur activité IAAS cloud est incertain. Ils ne doivent donc être choisis qu'après avoir soigneusement soupesé tous les risques qu'ils présentent.

Le "Magic Quadrant" de Gartner : positionnement des solutions d'infrastructure "en tant que service" IAAS dans le monde en avril 2018



Les **Acteurs** de niche sont d'excellents fournisseurs pour les cas d'usages dans lesquels ils se spécialisent.

» Amazon, IBM, Google, Microsoft et Alibaba, enrichissent leur panel d'offre de fonctionnalités IA disponible dans le cloud, sous forme de catalogue de briques, disponibles "sur étagère" et prêtes à l'emploi.

Quel choix d'infrastructure pour accueillir l'IA ?

Cloud public, privé ou hybride : une adoption sur-mesure

Le cloud public

— Le **cloud** est dit **"public"*** lorsque les ressources cloud (serveurs, stockage etc.) sont mises à disposition par un fournisseur tiers et livrées via Internet. Toute l'infrastructure reste la propriété du fournisseur de solution cloud. Les ressources sont mutualisées entre plusieurs organisations tout en assurant l'étanchéité entre chacune d'elles.

Exemple d'utilisation :

Applications bureautique, environnement de déploiement et de tests...

Avantages :

Païement à l'usage du service (sur des échelles de temps très courtes), absence de maintenance (responsabilité du fournisseur), extensibilité des ressources à la demande quasi-illimitée.

Le cloud privé

— Le **cloud** est dit **"privé"*** lorsque les ressources cloud (serveurs, stockage etc.) sont utilisées de façon exclusive par une organisation, via un réseau privé ou internet. Il peut être situé physiquement dans l'organisation (cloud interne) ou hébergé par un fournisseur tiers. Si toute l'infrastructure n'appartient pas à l'organisation, la maintenance des services et de l'infrastructure est de sa responsabilité. Par rapport au cloud public, le cloud privé est personnalisable : il s'adapte aux besoins.

Exemple d'utilisation :

Dans le cadre d'un besoin de contrôle optimal de l'environnement.

Avantages :

Flexibilité de l'environnement (personnalisation), sécurité accrue (niveaux de contrôle et de sécurité additionnels possibles du fait de la non mutualisation).

Le cloud hybride

— Le **cloud** est dit **"hybride"*** ou **"meilleurs des deux mondes"** lorsqu'il combine cloud public et cloud privé (ou infrastructure locale "on premise").

Exemple d'utilisation :

Du cloud public pour des besoins ayant trait à des volumes importants et moins exigeants en matière de sécurité (ex : applications de bureautique) et du cloud privé (ou une autre infrastructure locale), pour des opérations sensibles et stratégiques (ex : données financières).

Avantages :

Contrôle complet sur les ressources sensibles, flexibilité (tirer parti de ressources supplémentaires uniquement lorsque c'est nécessaire), facilité de migration vers le cloud (migration graduelle des charges).

» **Dans les entreprises, le choix d'architecture* privé, publique ou hybride (mixte publique et privée) s'évalue en fonction de 4 facteurs : le niveau de sécurité, les exigences technologiques, l'agilité dans les ressources et les compétences requises.** (cf. schéma [Quelles sont les principales raisons de choisir p.70](#))

Quelles sont les principales raisons de choisir :



Source : Artefact

Devenir agnostique des solutions cloud pour maximiser la valeur de l'IA

La variété des **couches d'abstraction*** et la gradation de l'adoption du cloud multiplient le champ des possibles pour les organisations qui adaptent leurs demandes en services cloud de plus en plus finement.

Le marché des fournisseurs, en forte croissance et concurrentiel, incrémente cette multitude d'options avec des services de plus en plus spécifiques sur des aires d'expertises pointues. Par exemple ils fournissent des accès à des fonctionnalités précises d'intelligence artificielle, reconnaissance d'image (API Vision de GCP, **Google Cloud Vision***), reconnaissance vocale (Amazon Transcribe)...

Cette multitude de possibilités facilite l'adoption de l'IA, permettant aux organisations de choisir le "meilleur des mondes" en combinant les solutions, pour obtenir un service le plus adapté possible à leurs besoins.

Ce choix initial peut cependant amener à évoluer en fonction de nouveaux besoins ou de nouvelles équations de rentabilité. Se pose donc la question de la dépendance aux solutions adoptées.

L'essor de la portabilité, un nouvel élan

La portabilité des données permet aux utilisateurs du cloud de récupérer leurs données auprès de leurs fournisseurs de solution cloud. Elle a vocation à renforcer le contrôle qu'ils ont sur leurs données : connaissance du volume et du type de données disponibles, demande de suppression... Ce principe est un point fort du RGPD et du projet de marché unique numérique européen.

Facebook, Google, Microsoft et Twitter ont lancé en 2007 le projet DTP (Data Transfer Project), afin d'aller au-delà de la simple prise de connaissance des données et de pouvoir proposer un "cloud switch" permettant de transférer directement les données d'un fournisseur à l'autre.

Portabilité et interopérabilité des données permettent donc d'assurer une plus grande indépendance aux solutions cloud, et de lever une barrière à l'entrée pour l'usage de briques IA de leurs catalogues.

01

Le point clé de l'adoption de l'IA dans les organisations est celui de la recherche de la valeur.

L'IA est d'autant plus complexe à mettre en oeuvre que son impact est imprévisible.

En bref.

A titre de comparaison, le digital a ceci de plus simple que les solutions digitales (e-commerce, site institutionnel, dématérialisation, etc.) sont connues avec un enjeu essentiellement de maîtrise, d'exécution et de transformation d'entreprise.

Côté intelligence artificielle, **la preuve du potentiel cas d'usage par cas d'usage est une étape indispensable avant d'entamer le grand défi de la transformation.**

02

Scale AI

Bâtir de solides fondations pour faire de l'impact à l'échelle

02

Grâce aux nouvelles possibilités techniques liées au cloud, l'intelligence artificielle permet la création de nouveaux produits et services, et représente un avantage concurrentiel stratégique pour les entreprises.

Selon une étude réalisée par Bain & Company en 2017, les prises de décisions sont deux fois plus orientées par la donnée et cinq fois plus rapides grâce à l'IA.

Cependant le passage à l'échelle reste complexe :

31%

des dirigeants estiment qu'ils n'ont pas les bonnes ressources en interne pour se lancer dans une stratégie IA¹

66%

d'entre eux estiment que les données ne sont pas disponibles et que les outils ne sont pas adaptés à leur projet²

70%

des entreprises multiplient les POCs sans passage en industrialisation²

Intro.

L'enjeu est de lutter contre l'éparpillement et regrouper les efforts pour faire de l'impact à l'échelle. Cela passe par la construction de solides fondations techniques et humaines, sur lesquelles l'entreprise puisse se reposer pour développer sa stratégie.

Ce chantier nécessite le lancement de 3 grands programmes :

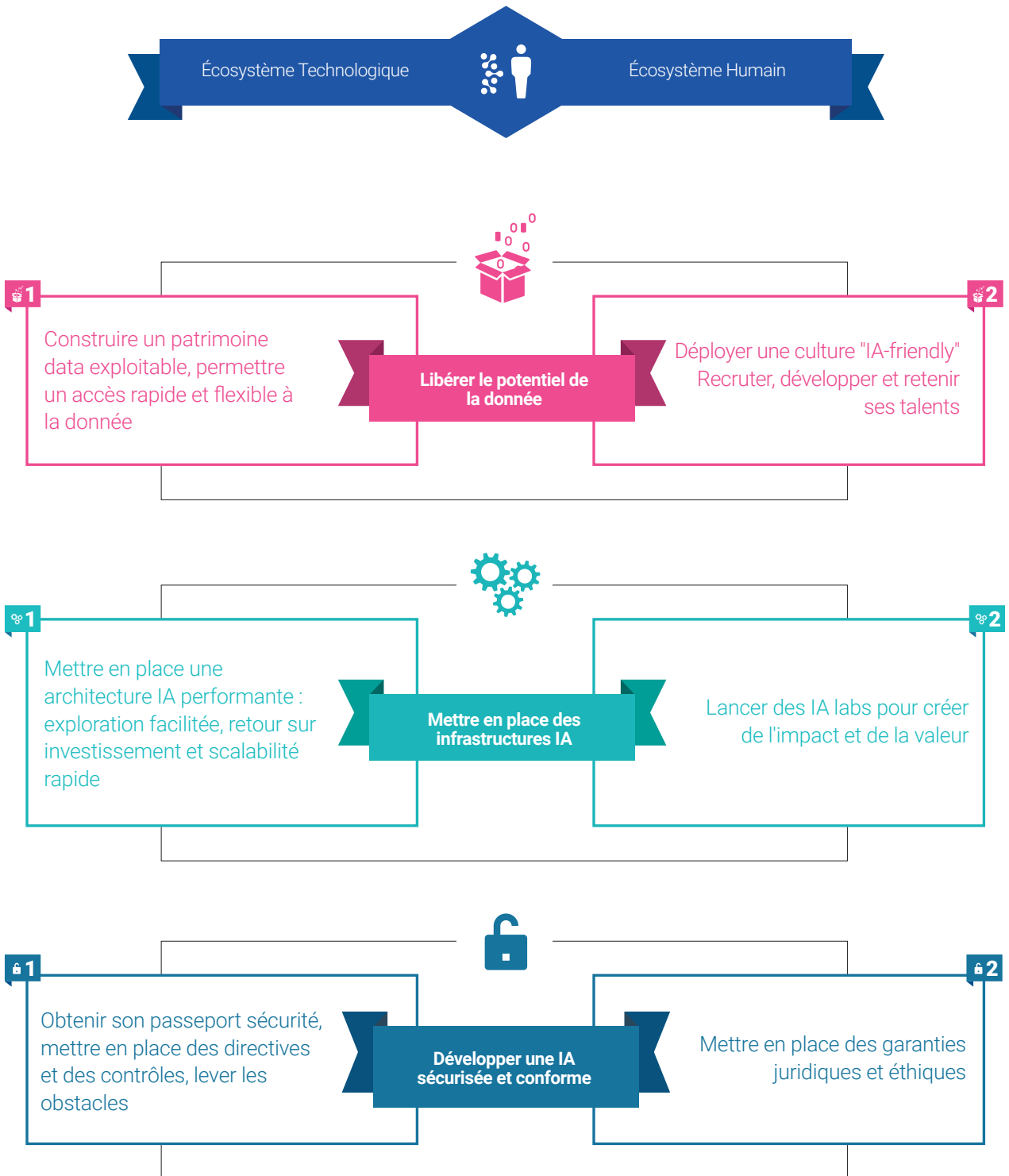
- Libérer la donnée,
- Mettre en place les bons enablers, techniques et humains,
- Développer une IA conforme, éthique et sécurisée.

Aujourd'hui, au regard de l'envergure des sujets, des enjeux associés et des budgets investis, l'IA mérite que l'on change d'échelle et de paradigme.

¹ Etude sur l'utilisation de l'intelligence artificielle au sein des entreprises françaises, Réalisée par Tata Consultancy Services et IDC

² Etude Boston Consulting Group/MIT Sloan Management Review, 2017

Les fondations de l'IA@Scale



Source : Artefact

Libérer la donnée

Construire un patrimoine data exploitable, permettre un accès rapide et flexible à la donnée

Une donnée disponible : fournir un accès centralisé à des données décentralisées

La tendance depuis quelques années est à la rationalisation des plateformes et à la centralisation des données au sein de datalakes permettant le croisement et la modélisation de grands volumes de données déstructurées.

Ce principe de centralisation de l'intégralité des données de l'entreprise au sein d'une même plateforme est souvent difficile à opérationnaliser au regard de la complexité des infrastructures et contraintes des directions IT (par exemple certaines contraintes réglementaires nécessitent d'isoler certains gisements de données).

» **L'enjeu majeur n'est pas celui de la centralisation des données, mais celui de la mise en place d'accès centralisés à des données pouvant rester décentralisées.**

Par le biais d'une interface, des systèmes de requêtes distribuées permettent d'exposer et traiter des données provenant de gisements divers, évitant ainsi les sujets complexes de gouvernance de données associés à la centralisation du patrimoine data.

Un accès centralisé à des données décentralisées

Par exemple, Presto permet d'écrire des requêtes SQL qui référencent des données éclatées dans différents systèmes, sans les dupliquer

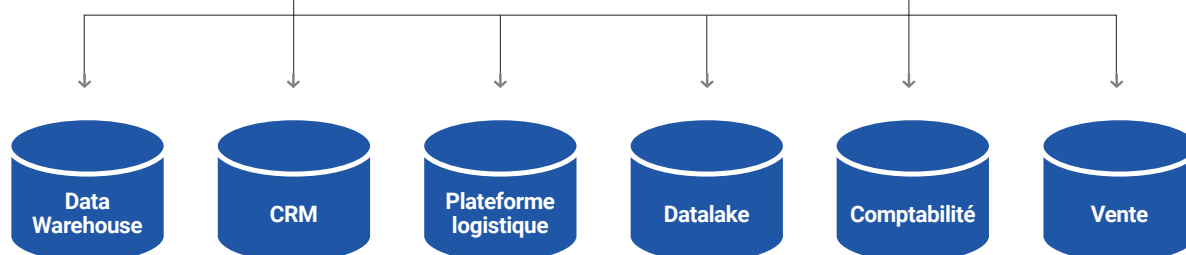
Requête SQL

presto

Requête API

GraphQL

GraphQL expose les données en tant que modèle unique via une API en récupérant et en combinant des données provenant de sources multiples



Source : Artefact

Une donnée compréhensible : fournir un accès centralisé à des connaissances décentralisées

La connaissance de la donnée est souvent centralisée chez un groupe restreint de personnes, sans être correctement documentée. Les conséquences sur le projet et sur l'entreprise sont alors multiples :

Coût d'opportunité

Par manque de connaissance, on restreint l'apprentissage de l'algorithme à des données connues (certaines données pertinentes mais non connues de l'équipe qui réalise le cas d'usage sont de fait exclues).

Perte de productivité

L'équipe est dépendante d'une tierce personne dans l'exploitation de la donnée, ce qui retarde souvent l'avancement des projets.

» **Une fois le patrimoine data de l'entreprise construit, il est stratégique d'en démocratiser la compréhension et l'usage via la mise en place de dictionnaires de la donnée répertoriant et explicitant l'ensemble des données disponibles (typologies, gisement source, significations technique & métier etc..).**

L'objectif final est de réduire les dépendances et rendre autonome les ressources (souvent des ressources externes ou de récentes recrues n'ayant pas encore acquis de connaissances tangibles sur la structure et signification de la donnée dans l'entreprise).

Une donnée exploitable : définir et lancer des stratégies de Quality-by-Design

Utiliser la qualité de la donnée pour se différencier

L'importance de la qualité de données - sujet déjà clé à l'ère du big data - se retrouve aujourd'hui renforcée avec l'intelligence artificielle.

"La mise en place d'une stratégie de qualité de données est un pré-requis au lancement d'une stratégie d'IA."

Clément Cardi,
Cloud Solution Engineer, Google

Toutes les entreprises ayant accès aux mêmes algorithmes disponibles sur l'étagère, le principal facteur de différenciation se situe au niveau de la matière première qui leur est fournie : la donnée.

Le machine learning est basé sur la notion d'apprentissage ; il ne peut y avoir d'apprentissage juste sur des données inexistantes ou erronées.

Pour Clément Cardi, Cloud Solution Engineer chez Google, "La mise en place d'une stratégie de qualité de données est un pré-requis au lancement d'une stratégie d'intelligence artificielle".

Lancer des programmes de "Quality-by-Design"

Alors que doit faire une entreprise pour tirer le meilleur parti de l'intelligence artificielle ?

Face à la complexité de la tâche et à l'urgence du sujet, il est souvent difficile de savoir par où commencer.

» **Le point clé est de privilégier des actions concrètes sur des périmètres de données critiques mais restreints.**

Kit de lancement

Définir une stratégie de qualité de donnée :

- + Définition des critères de qualité de donnée (ex : la complétude de la donnée ainsi que sa fraîcheur sont deux critères clés)
- + Choix de deux KPIs à prioriser par critère de qualité (ex : le taux de complétude de la donnée)
- + Définition et priorisation des périmètres de données sur lesquels lancer la stratégie (commencer par les périmètres coeur business, qui vont être les plus sollicités dans le cadre de projets d'intelligence artificielle)

Définir les modalités de mise en oeuvre opérationnelle de la stratégie

- + Définition des seuils de qualité par KPI (ex : le taux de complétude de la donnée doit être de 90% minimum)
- + Choix du type de monitoring à effectuer (ex : système de notifications, dashboard de qualité...)
- + Définir une stratégie de communication autour de la qualité des données (ex : comment prévenir les utilisateurs, à quelle fréquence etc.)

Mettre en place une gouvernance autour de la qualité des données

- + Définition de rôles spécifiques (ex : responsable de sources)

Construire un patrimoine data accessible, compréhensible et exploitable est le premier chantier de notre programme "Libérer la donnée".

L'avalanche de données engendrée par le big data n'a de sens que si les informations produites peuvent être facilement appréhendées et interprétées.

L'acculturation de tous, la maîtrise de la façon dont les données sont produites puis exploitées est un des grands enjeux des entreprises.

Garder l'humain au coeur de la transformation

Créer et diffuser une culture data et IA

Pour tous : Démystifier, "IAAlphabétiser" et Engager

Alors que la responsabilité d'implémenter des nouvelles technologies appartient au management, les opérationnels - d'ordinaire exclus des choix technologiques - sont les premiers à être impactés : disparition de certains métiers, modification du contenu d'autres, apparition de nouveaux besoins en formation... l'intelligence artificielle déstabilise les dispositifs en place.

» Les programmes d'acculturation sont les premiers catalyseurs des transformations IA en entreprise.

En comprenant les impacts de la transformation, chacun est alors davantage susceptible de s'y intégrer.

Programme d'acculturation

"Adopte l'IA"

Démystifier : rassurer les participants sur les impacts de l'intelligence artificielle et des nouvelles technologies

- + Être transparent sur les ambitions de l'organisation, en communiquant a priori sur les enjeux et impacts de la transformation, avec un langage non technique, compréhensible par tous.
- + Favoriser et multiplier les échanges entre experts et néophytes via l'organisation d'évènements rassembleurs.

Alphabétiser : donner à chacun les armes pour comprendre l'IA et ses enjeux

- + Dispenser des séances d'acculturation à la data et à l'intelligence artificielle ouvertes à tous, touchant du doigt un éventail d'exemples et d'enjeux. Privilégier l'étendue des sujets à la profondeur.
- + Développer un langage commun, en démocratisant le "jargon technique" via la multiplication de supports pédagogiques qui favorisent les échanges et la compréhension entre experts et néophytes.

Engager : impulser une transformation inclusive, où chacun a un rôle à jouer

- + Embarquer l'ensemble des fonctions support dans la transformation (Ressources Humaines, Finance, Juridique...).
- + Communiquer sur les projets et leurs impacts. Privilégier les dispositifs de communication fil rouge aux communications ad hoc.

Pour les opérationnels : recentrer les efforts sur la créativité

L'humain va devoir apprendre à travailler avec la machine. Pour la plupart des métiers, l'intelligence artificielle va "augmenter l'humain" et non le remplacer, elle va automatiser les tâches chronophages à faible valeur ajoutée. David Kenny, manager général chez IBM, en est convaincu : "La digitalisation des tâches de base va permettre aux humains de se concentrer sur leur créativité. Ils pourront ainsi augmenter leurs compétences et leur accès au savoir".

» **Même si un débat existe sur qui prendra l'ascendant - humain ou machine - sur les activités cognitives complexes, il est aujourd'hui admis que les entreprises se différencieront sur leurs capacités à innover, à créer.**

Par la voie d'ateliers d'idéation, le rôle des équipes du Chief AI Officer est d'accompagner l'ensemble des métiers sur l'identification de poches de valeur, ainsi que sur la génération d'idées et de solutions pour répondre à une problématique donnée. Parmi les méthodes d'idéation collective, la méthodologie **Design Thinking*** est la plus couramment utilisée.

"La digitalisation des tâches de base va permettre aux humains de se concentrer sur leur créativité."

David Kenny,
Manager Général, IBM

Pour les experts de la donnée et les équipes dirigeantes : faire évoluer les mentalités

Apprendre à optimiser des boîtes noires

La transformation IA nécessite un changement d'état d'esprit et de paradigme pour les spécialistes de la donnée (data miners, développeurs, data scientists etc.).

On passe d'un environnement maîtrisé, où chaque étape est pensée et construite par l'homme, à une collaboration Homme-machine où les objectifs sont définis en amont, mais où les méthodes de résolution sont bâties et optimisées par des algorithmes dont la complexité n'est compréhensible que par un nombre restreint d'experts.

Les mêmes qui demandaient à un ordinateur d'exécuter une série de règles définies par leur soin, vont devoir aujourd'hui faire confiance aux algorithmes qui sont par nature des "boîtes noires".

» **La différence majeure ne se fait plus au niveau de la programmation du modèle, mais au niveau de la préparation et mise en qualité de la donnée, et du choix de l'infrastructure technique.**

Les équipes data doivent se concentrer sur la construction d'infrastructures agiles et scalables, assez puissantes pour permettre la mise en oeuvre de cas d'usages IA innovants à moindre coût.

Penser "Re-use" et "Open source"

Pour Clément Cardi, Cloud Solution Engineer chez Google, il est vital pour les équipes techniques d'instaurer une culture de développement pragmatique, qui capitalise sur l'open source. "Il faut arrêter de vouloir tout réinventer. L'open source est un travail qui est fait une fois et qui peut être réutilisé à souhait. Aujourd'hui la plupart des data scientists en France utilisent des modèles créés par des chercheurs mathématiciens, et les adaptent à leurs problèmes spécifiques".

Ces pratiques ont la vertu de mettre l'intelligence artificielle à la portée de tous les métiers de la donnée, qui peuvent appliquer à leur problème des modèles de machine learning trouvés sur l'étagère. Toujours pour Clément Cardi, "À la cible, l'objectif serait de ne plus être dépendant de la data science, d'avoir uniquement besoin d'un développeur chargé d'intégrer le service IA dans les systèmes d'information. L'IA doit être accessible à tous. Demain l'idéal serait de se reposer ponctuellement sur les data scientists pour la création de modèles complexes et sur-mesure, assez malléables pour être retravaillés par des profils aux compétences data moins avancées".

Penser Cloud

En plus d'un nouveau paradigme data, les différents acteurs sont tenus de se familiariser avec un nouveau paradigme outils. Pour Valérie Coscas d'Amazon, "il faut penser cloud. Le cloud est un état d'esprit". Ce nécessaire changement de philosophie doit toucher l'ensemble des fonctions de l'entreprise et des cadres dirigeants. Toujours selon Valérie Coscas, "Pour que la transformation soit inscrite dans la

durée, le bouleversement des processus doit être généralisé (changement des processus d'achat, mise en place d'une nouvelle gouvernance...). Si l'on prend l'exemple du CFO, le cloud transforme la structure de coûts des investissements IT (OPEX vs CAPEX), c'est un changement total de paradigme pour lui et ses équipes".

"Demain, l'idéal serait de se reposer ponctuellement sur les data scientists pour la création de modèles complexes et sur-mesure, assez malléables pour être retravaillés par des profils aux compétences data moins avancées."

Clément Cardi,
Cloud Solution Engineer, Google

Recruter, développer et retenir les talents

Recruter une équipe pluridisciplinaire

Le mariage entre technique et métier au sein de l'équipe cœur

L'IA s'inscrit en opposition à la séparation des compétences techniques et business, auparavant observées dans la gestion des projets SI. Ces projets sont traditionnellement structurés autour d'un maître d'ouvrage (MOA), en charge

de rédiger l'expression des besoins et souvent déconnecté des contraintes techniques ; et d'un maître d'œuvre (MOE), responsable de mettre en oeuvre le cahier des charges et d'ordinaire très éloigné des contraintes métiers.

Même si cette dichotomie a tendance à s'estomper ces dernières années sous la pression de la transformation digitale et data, on observe aujourd'hui encore dans la plupart des entreprises une gestion en silos des compétences business et techniques qui ne parlent pas le même langage.

La transformation IA a vocation à transformer le cœur business des organisations ; pour tenir sa promesse elle ne peut se passer des nouvelles technologies cloud.

» **Le changement doit être piloté par une équipe cœur rassemblant des profils techniques et business, parlant le même langage et œuvrant pour un même objectif.**

Sans profil business à bord, le risque de livrer un produit IA déconnecté des besoins métiers et rognant sur la promesse initiale est accru ; sans profil technique, le projet ne peut être déployé à l'échelle.

— L'AI Product Owner : le sweet spot

Au sein de l'équipe cœur, le business est incarné par l'AI Product Owner qui est responsable du planning, de la livraison finale ainsi que de la traduction du brief métier et des attentes sponsors en spécifications techniques. "Trilingue" SI/ Data/Business, il oscille entre vérification des livrables techniques et évaluation de la faisabilité des demandes métiers. Il est en capacité de challenger les équipes techniques sur leurs choix et participe au design du produit final.

Le changement doit être pensé par une **équipe cœur** rassemblant des profils techniques et business, parlant le même langage et œuvrant pour un même objectif

Son rôle est crucial tout au long du projet : choix des données à intégrer dans le modèle, pondération de l'importance de chaque type de données, évaluation des résultats etc. Pour Edouard de Mezerac, Partner chez Artefact, ces profils sont les plus difficiles à acquérir pour une entreprise : *"Les talents les plus rares ne sont pas les data scientists mais les talents*



Edouard de Mezerac
Partner chez Artefact

hybrides qui sont capables de réunir dans le même cerveau une compréhension de la problématique business et des techniques data et IA pour les résoudre. C'est le sweet spot".

En 2017, le gouvernement chinois a fait de l'IA sa priorité en rendant public le plan de développement national qu'il souhaite mettre en place pour l'émergence d'une intelligence artificielle "made in China". En réponse, JD.com, un des deux plus importants e-retailers chinois a embauché 800 data scientists. Un an plus tard aucun cas d'usages majeur n'était passé en production. Le post mortem fait le constat du manque de product owners au sein de l'équipe coeur.

"Les talents les plus rares ne sont pas les data scientists mais les talents hybrides qui sont capables de réunir dans le même cerveau une compréhension de la problématique business et des techniques data et IA pour les résoudre. C'est le sweet spot."

Edouard de Mezerac,
Partner chez Artefact

— Data scientists, data engineers et data architects : le trio gagnant côté technique

Alors que le data scientist voit la donnée comme de l'information, le data engineer la voit comme de la matière.

Le data engineer travaille en amont et en aval du projet IA, collecte et traite la donnée issue du datalake et déploie en fin de projet le modèle en production. Il est responsable de la mise en place et du maintien des infrastructures.

» **Le rôle du data engineer est central dans la réalisation d'un projet IA, et se voit renforcé dans le cadre de projets réalisés sur le cloud.**

Les data engineers mettent en place les flux de données qui nourrissent les analyses et les modèles des datascientists. Ils permettent à ces derniers de se concentrer sur le nettoyage, la préparation et la modélisation de la donnée.

Selon Douglas Willcocks "Le data scientist doit s'abstraire complètement des problématiques d'infrastructure et se concentrer sur le déploiement d'un modèle optimal,

augmentant le nombre d'itérations et donc par conséquent la qualité du résultat final".

Contrairement au data scientist et au data engineer, le data architect est plus dans la conception que dans la réalisation du projet.

» **Le data architect définit l'architecture cible en cohérence avec la stratégie IT et anticipe l'intégration du produit final au SI legacy.**

Son rôle est clé dans la conception d'architectures hybrides cloud / on premise.

Dans les directives de recrutement des profils techniques, les compétences techniques, 'hard skills', sont évaluées au même niveau que d'autres "plus douces", par exemple la capacité à vulgariser des sujets techniques complexes ou encore l'appétence business.

» **Interdisciplinarité et collaborativité sont clés pour délivrer des projets d'IA à l'échelle, par définition transverses aux organisations.**

Le soutien des experts dans l'équipe étendue

— Des informations terrain grâce aux équipes opérationnelles

En plus d'un Business Owner porteur du projet côté métier, les équipes opérationnelles sont également parties prenantes pour apporter des informations terrains orientant le paramétrage du modèle et valider avec l'AI Product Owner la pertinence de la solution finale.

» **Plus l'expertise métier est pointue, plus le rôle des acteurs terrain est important.**

Par exemple, un acteur du secteur des Télécoms s'est lancé en 2017 dans la mise en place d'un algorithme de machine learning prédisant les scénarii optimaux de déploiement de la fibre en France. Sans l'aide des concepteurs réseaux, fins connaisseurs des spécificités terrains, le projet n'aurait pu aboutir.

— Des juristes formés aux problématiques IA et Data

La transformation IA révèle de nouveaux profils. Pour Rémi Cauchois, juriste chez Orange France, le juriste "privacy" est un responsable juridique doté d'une appétence sécurité, qui dispose d'une connaissance fine des gisements de données, des technologies et des usages métiers associés. La maîtrise de la complexité opérationnelle des projets d'intelligence artificielle est un pré-requis à la maîtrise de la complexité juridique.

Le rôle du juriste privacy est de garantir la conformité des traitements réalisés, tout en considérant l'incertitude inhérente à tout projet d'IA (à cause de l'effet boîte noire des algorithmes).

— Des équipes SI orientées service

Dans le cadre d'une architecture full cloud, le SI est libéré des tâches de gestion et maintenance des infrastructures. Son rôle est de fournir des services à ses clients internes. Ces services peuvent être offerts par les infrastructures on premise, par le cloud, ou de manière hybride, cette dernière option rendant plus ardu le travail d'intégration. Ces équipes doivent faire partie de l'équipe étendue, pour réfléchir à l'intégration harmonieuse du produit final aux infrastructures existantes.

"Alors que le temps des POCs était l'apanage des métiers, le temps de l'IA à l'échelle est celui du SI"

Vincent Luciani,
Co-fondateur et Global COO d'Artefact

Chez Allianz, chaque cas d'usage IA est développé dans une optique industrielle qui s'exécute sur les contraintes du SI. Dès l'initiation du projet cloud, la réflexion de son intégration au SI legacy est entamée.

Pour Steve Cheney, responsable de l'équipe data engineer chez Allianz, "Le rôle premier du SI est de rendre service à son client, quelque soit le type d'infrastructure choisie pour réaliser le cas d'usage. Le service du SI est agnostique de toute plateforme technique".

Le sponsoring des change leaders

— Tout le monde à bord

Afin de mettre l'IA au coeur des grands enjeux et transformations de l'entreprise, pour Valérie Coscas d'Amazon, "Il est nécessaire d'avoir le sponsoring de l'ensemble du COMEX, chaque membre étant de prêt ou de loin impacté par la transformation IA".

L'erreur serait de penser que les sponsors de la transformation sont uniquement les directeurs techniques ou SI (CTO).

» **La transformation doit être impulsée au plus haut niveau hiérarchique, par le CEO qui met en place et choisit son équipe pluridisciplinaire de "change leaders" (membres du COMEX porteurs d'un chantier de la transformation).**

— Le CAIO, le capitaine du navire

Pour impulser la transformation, l'équipe dirigeante a besoin d'un expert de confiance, capable d'identifier les opportunités de nouvelles sources de revenus ou d'économies grâce à l'IA.

Pour Andrew Ng, le fondateur de Coursera, "le Chief AI officer possède la double expertise technique/business, permettant de capitaliser sur les nouvelles technologies IA toujours au service de considérations cœur business".

Le CAIO peut dans certaines organisations être représenté par le Chief Data Officer ou le Chief Digital Officer.

Le rôle du CAIO est également d'insuffler la transformation au sein de l'équipe dirigeante. Il joue un rôle majeur dans l'acculturation des membres du COMEX et dissipe les inquiétudes sur les impacts financiers de l'IA. Beaucoup d'organisations échouent à capitaliser sur l'IA car les dirigeants ne comprennent pas ou ne connaissent pas l'étendue des possibilités offertes par la discipline.

Embaucher un CAIO est un signal très positif envoyé au marché de l'emploi qui va permettre d'attirer de nouveaux talents.

— Le Chief People Officer, le bras droit du CAIO

Le rôle du CPO ne se limite pas à l'anticipation de l'hypercroissance. Son rôle est d'assurer que l'organisation attire, recrute, développe et retienne les talents.

Le CPO comprend la stratégie IA de l'entreprise et la connecte à sa propre stratégie de recrutement et de formation. Pour Lewis Lamonica, Principal à True Search *"Si votre prochain challenge est le traitement automatique du langage, alors le rôle de votre CPO est d'aider le leadership à réfléchir à la manière dont cela influe sur la stratégie de l'entreprise en matière de recrutement, d'intégration, de développement et de rétention des talents et de construire un plan pour atteindre l'objectif"*.

Le sens du business et des affaires est le facteur différenciant pour ce type de profils. Si l'on prend l'exemple des GAFAs, les candidats retenus sont ceux disposant d'une expertise business forte, extérieurs au monde des Ressources Humaines (par ex : Lori Goler, embauchée chez Facebook et issue du secteur du marketing ; Eileen Naughton, recrutée chez Google alors qu'elle provient du secteur commercial).

Développer et retenir les talents

Investir dans la formation

Pour Guillaume De Roquemaurel, CEO d'Artefact, le point clé est de s'assurer d'avoir les bonnes personnes à bord et se focaliser sur leur montée en compétence, *"il faut quitter le sujet pur technologique, il faut revenir à l'humain et à l'opérationnel"*.

Steve Cheney, met également l'accent sur l'enjeu de formation : *"les moteurs IA sont souvent spécifiques aux fournisseurs de cloud et supposent une véritable formation. Pour répondre à cet enjeu, Allianz contractualise avec les éditeurs de solutions pour apporter le niveau d'accompagnement nécessaire à la maîtrise de ces nouvelles technologies"*.

Pour Valérie Coscas, les investissements sont là : *"nos clients investissent massivement dans la formation de leurs équipes, et ce à plusieurs niveaux : certification des équipes techniques sur les technologies cloud, trainings des directeurs sur la philosophie et les enjeux du cloud, organisation d'atelier d'idéation..."*.

Pour y parvenir, les départements Ressources Humaines doivent être intégrés tôt dans la transformation, en participant notamment aux sessions pilotes des ateliers d'acculturation. Leur mission requiert d'avoir une vision claire des enjeux et besoins pour décliner les plans de formation adéquats, adaptés à la multiplicité des rôles et compétences nécessaires à la mise en place d'un programme d'intelligence artificielle à l'échelle.

Ces programmes de montée en compétence se concentrent sur deux axes :

- + Le repositionnement et l'élargissement des compétences des profils data existants
- + Le développement de la créativité et de "l'out of the box thinking" pour les profils plus opérationnels.

Guillaume de Roquemaurel
CEO d'Artefact



Fidéliser : le défi des Chief People Officers

Une des missions clé du CPO et de l'ensemble des équipes Ressources Humaines est de développer les conditions de travail idéales, qui permettent à ces profils de s'épanouir et de développer leurs compétences au bénéfice de l'entreprise.

Intérêt, diversité et multiplicité des sujets traités, travail en équipe pluridisciplinaire, présence d'un "lead tech" favorisant le développement des compétences techniques... sont autant d'aspects sur lesquels le CPO va devoir focaliser son attention pour retenir ses talents.

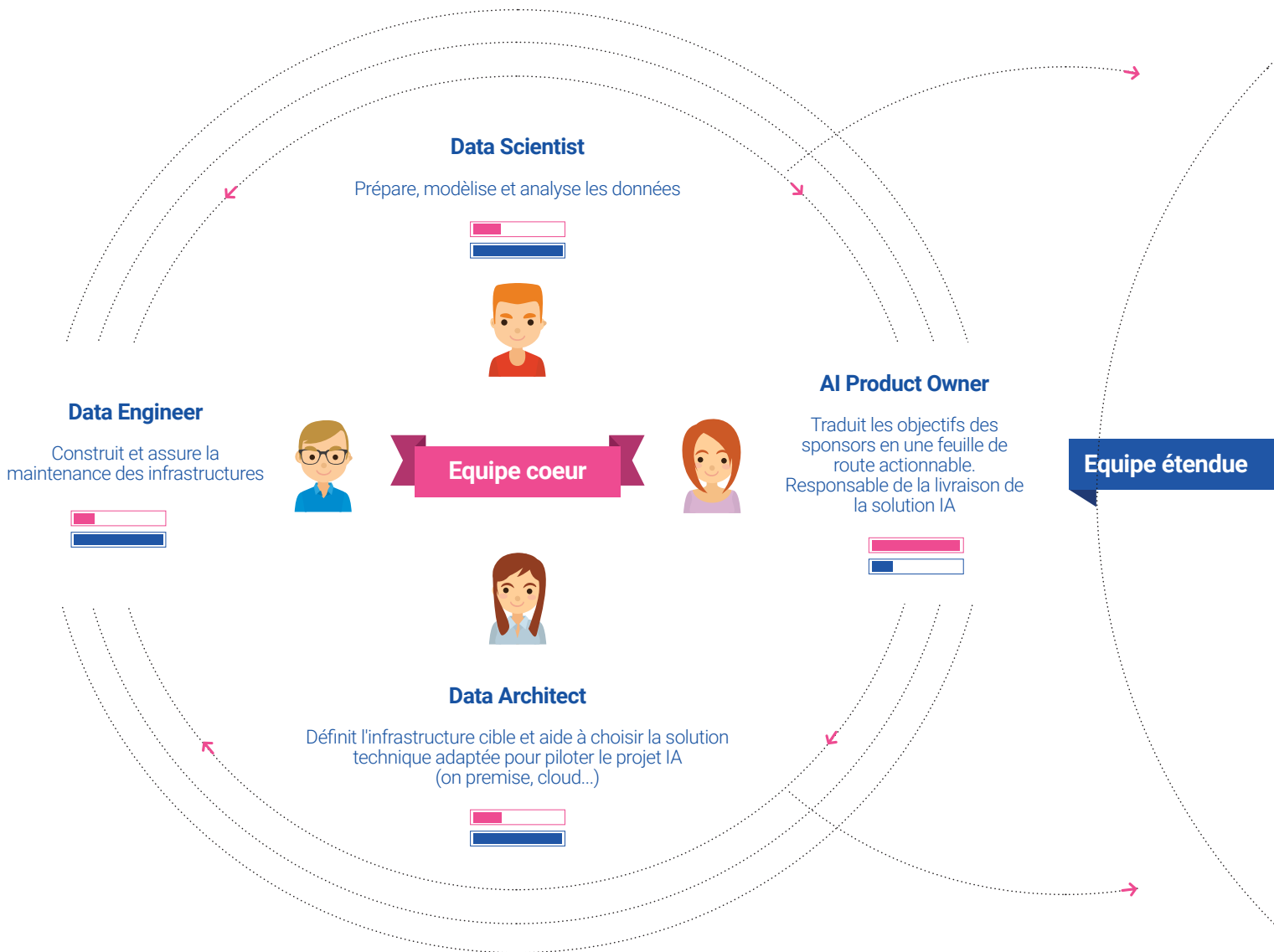
Une fois la donnée libérée grâce à la mise en place

- + d'un patrimoine data accessible, compréhensible par tous et exploitable
- + d'une organisation 'IA driven' qui IAlphabétise, recrute, développe et retient les talents,

Il s'agit d'être en capacité de générer un maximum d'impact et de valeur.

L'exploitation optimale de la donnée ne peut se faire sans la mise en place d'enablers à la fois techniques et humains, permettant d'accueillir l'IA au coeur de l'entreprise.

Les rôles clés pour la réussite des projets IA@Scale



 % Compétences métiers
 % Compétences techniques
 Source : Artefact



BUSINESS OWNER

MARKETING, VENTES, RH, SERVICE CLIENT, JURIDIQUE...

Client interne du projet IA :
Identifie les objectifs & KPIs du projet



EXPERTS JURIDIQUE & SÉCURITÉ

Valide la conformité légale de la solution IA et des outils choisis.

Responsable de la sécurité et de la confidentialité des données et infrastructures



EXPERTS IA & DATA

Responsable des données et des infrastructures



ÉQUIPE METIER OPÉRATIONNELLE

Fournit les insights utilisateurs et apporte une validation business

Change leaders IA



CEO

Conduit et orchestre l'ensemble de la transformation IA



CTO Chief Technology Officer

Responsable de la vision globale de l'infrastructure et des technologies du cloud



CAIO Chief AI Officer (ou CMO, ou CDO...)

Leader de la transformation IA. Traduit la vision IA du CEO en une feuille de route actionnable



DPO Data Protection Officer

Responsable de la sécurité, de la confidentialité et de la gouvernance des données à caractère sensible



CPO Chief People Officer

Harmonise le recrutement et la gestion des talents en lien avec la transformation IA de l'entreprise

Installer les bons enablers pour accueillir l'IA au coeur de l'entreprise

Mettre en place une infrastructure et des outils "AI friendly" : exploration facilitée, passage à l'échelle accéléré, rentabilité maximisée

Les 4 partis pris d'Artefact sur la manière d'optimiser ses infrastructures pour passer le cap de l'industrialisation

Tout projet d'intelligence artificielle est constitué d'une phase d'industrialisation et d'une phase de déploiement ou d'industrialisation.

L'industrialisation est le moment où la valeur de l'intelligence artificielle est capturée, où elle est déployée en tant que service consommé par d'autres utilisateurs ou systèmes. Si le coût de passage entre la partie conception et production est élevé, la bascule se fait rare, les équipes préférant alors itérer jusqu'à ce que l'incrément de valeur obtenu justifie le passage en production.

Comment favoriser l'industrialisation des projets d'intelligence artificielle ?

Simplifier vos infrastructures avec les microservices

► **Dans le cas d'applications complexes, il est nécessaire de découper la complexité en plusieurs unités fonctionnelles simples : les microservices.**

Chaque microservice fonctionne comme une brique de Lego qui évolue indépendamment des autres briques et permet une plus grande agilité de développement. Une brique peut être ajoutée, modifiée ou supprimée sans impacter l'ensemble de la structure.

Prenons l'exemple d'un data scientist qui crée un micro-modèle pour classifier une intention d'appel en call center ; il n'a pas besoin de l'intégrer à une application "mère"

pour en garantir le fonctionnement, le micro-modèle est un "mini expert" autonome dans la tâche qu'on lui demande d'exécuter.

Le data scientist peut par la suite développer un autre micro-modèle, proposant aux conseillers des arguments à énoncer en fonction de l'intention d'appel détectée. Ce second modèle pourra échanger par **APIs*** des informations avec le premier sans être contraint par son évolution.

Les deux micro-modèles ont été développés de façon disjointes, mais sont utilisés de manière composée grâce aux APIs. Chaque modèle peut de manière isolée être consommé par un autre service ou être partie d'une autre application, sans que l'autre en soit impacté.

Selon Douglas Willcocks, Global Chief Data Officer chez Artefact, "Les applicatifs d'intelligence artificielle sont des "data products" construits sous forme de microservices, dont l'enjeu principal pour les équipes projet est d'en maximiser la consommation. Plus la donnée est consommée, plus elle est génératrice de valeur. Le plus simple pour se lancer dans la construction de data products est d'utiliser des conteneurs".

Packager vos "data products" grâce aux conteneurs*

Les data scientists qui construisent les modèles n'ont pas l'expertise nécessaire pour gérer de manière optimale leur déploiement. Le data engineer entre en jeu, avec la responsabilité de rendre le travail du data scientist "éligible" à une mise en production. La spécificité de chaque modèle développé complexifie le travail du data engineer, qui doit s'entretenir avec le data scientist afin d'en comprendre les rouages.

» **L'enjeu est d'arriver à s'abstraire des éléments spécifiques et d'uniformiser la bascule entre prototypage et industrialisation.**

Les conteneurs permettent aux datascientists de "packager" l'intégralité de leur travail à l'intérieur d'un module à la géométrie connue (standards de poids, de taille, de composants etc.).

Ils proviennent d'un projet open source appelé Docker, devenu l'entreprise qui a lancé le format de conteneur le plus connu, rendant aujourd'hui les termes "dockers" et "conteneurs" interchangeables.

La personne en charge de déplacer le conteneur n'a pas besoin de savoir ce qu'il contient, le conteneur devient le standard. Le rôle du data engineer est de comprendre les pré-requis des conteneurs qui lui sont livrés (faut-il le faire tourner sur une machine TPU ? Comporte-t-il des données critiques nécessitant une mise en production sur des infrastructures on premise ?).



Douglas Willcocks
Global Chief Data Officer
chez Artefact

» **La mise en place de consignes standardisées via l'utilisation de conteneurs réduit le nombre d'interactions entre data engineer et data scientist, améliorant la scalabilité organisationnelle de l'entreprise.**

Réunifier vos infrastructures cloud et on premise grâce à Kubernetes

Un applicatif d'intelligence artificielle est composé de plusieurs conteneurs. Si on prend l'exemple d'un algorithme de machine learning, il est usuel de développer un conteneur pour le modèle, un second pour l'API qui l'expose au monde extérieur, et un troisième pour le ré-entraîner ponctuellement.

» **Un travail d'orchestration est nécessaire pour que les conteneurs tournent en même temps, qu'ils puissent communiquer entre eux, et qu'ils évoluent ensemble de manière cohérente lors d'une augmentation de charge.**

Kubernetes est une plateforme open source d'orchestration des conteneurs adoptée par la majorité des fournisseurs de cloud. Il s'agit d'une plateforme as-a-service que les data scientists utilisent pour déployer intuitivement leur modèle.

Elle est dotée d'une interface visuelle qui permet d'orchestrer la relation entre les différents conteneurs et fait en sorte que l'ensemble converge harmonieusement vers un monde idéal. Elle peut tout aussi bien déclencher et suivre l'exécution de règles pré-définies (exemple : la charge CPU de l'applicatif ne doit jamais passer au dessous de 50%) comme ajouter ou supprimer des conteneurs à la volée.

Un même conteneur peut aussi bien être déployé sur du cloud Amazon, Google, IBM, Microsoft, sur du cloud privé ou sur du on premise ; les conteneurs sont utilisables "Plug & Play" dans un écosystème IT où chaque option, chaque alternative a ses spécificités. Une application containerisée conçue sur un cloud public peut être migrée vers du cloud privé ou vers une architecture on premise.

Selon Douglas Willcocks "Avant, tout le monde disposait de machines physiques, et le principal enjeu était de découper ces machines physiques en machines virtuelles, de faire abstraction de la machine physique en elle-même, c'est ce qui a permis d'avoir des VM (Virtual Machines). Ensuite, avec l'arrivée du serverless, on a voulu s'abstraire des VM, mais on raisonnait toujours en tant que service provider. Aujourd'hui, on souhaite également faire abstraction du service provider en mettant une couche homogénéisante au dessus des data

centers on premise, au dessus des divers cloud privés et publics ; et de la même façon que vous ne choisissez pas la machine physique qui tourne dans votre datacenter, demain vous ne choisirez pas le cloud provider sur lequel tourne votre applicatif.”

» **Pour Douglas Willcocks, les entreprises sont maintenant libres de se concentrer sur 3 axes différenciants pour choisir l'hébergement de leurs applicatifs d'intelligence artificielle : le degré d'agilité, le coût et les garanties de sécurité offerts par les différentes solutions.**

Lors de son dernier Cloud Summit en Juin 2018, Google a annoncé le lancement de Kubernetes “on premise” dont le but est de répondre aux besoins des entreprises qui exécutent des microservices dans leur propre datacenter. Les machines tournent dans le data center on premise mais le travail de paramétrage, de monitoring et de maintenance est réalisé à distance par Google.

L'objectif est double :

- Proposer des services managés sur du on premise pour réduire la barrière à l'entrée (Kubernetes est une technologie complexe à maîtriser)
- Faciliter l'accès au cloud en utilisant Kubernetes comme une passerelle sécurisée pour déplacer progressivement certaines applications on premise vers le cloud

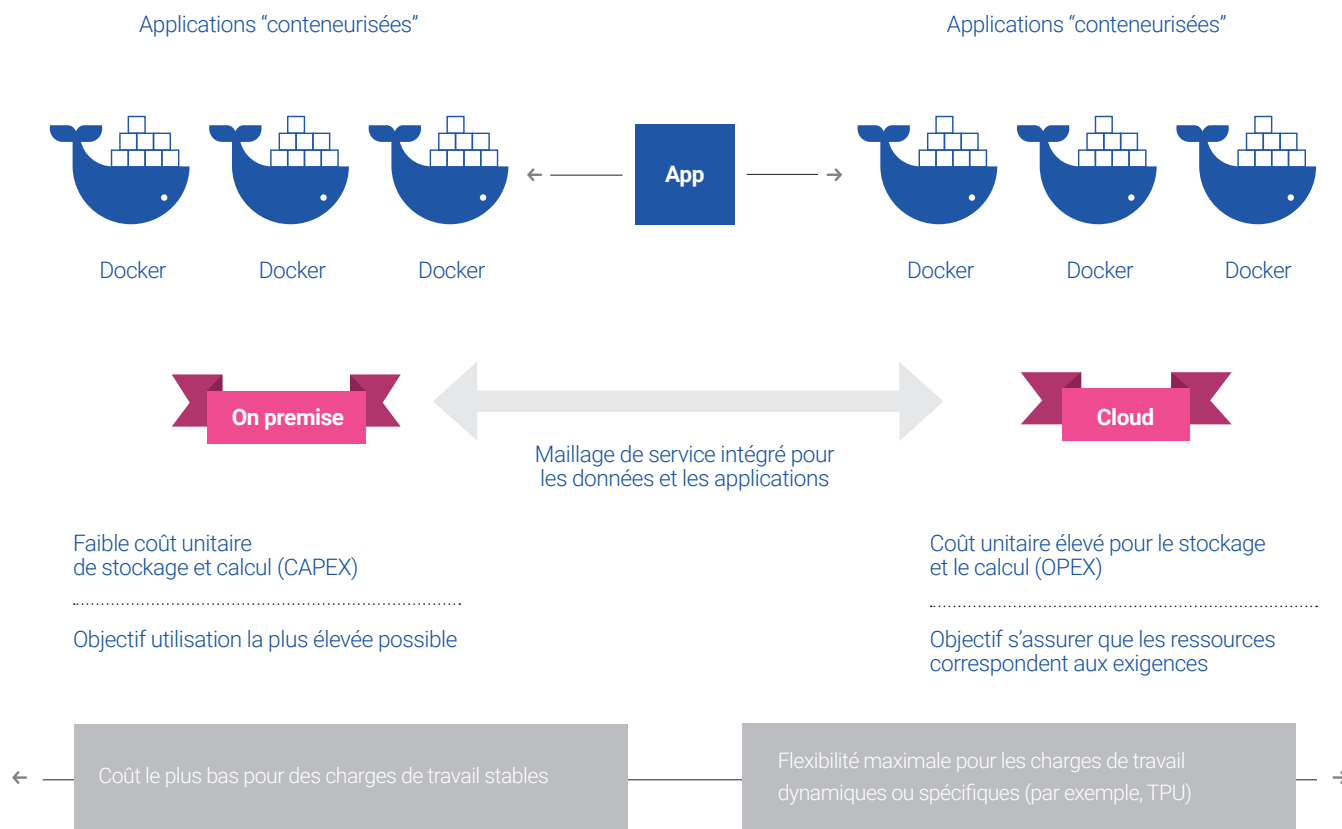
» **Grâce aux conteneurs et à leur orchestration, l'IA est devenue agnostique en termes d'infrastructures. Il n'y a plus besoin de choisir, de miser sur une plateforme ou une technologie. Cette portabilité garantie par Kubernetes permet aux entreprises de conserver leur souveraineté dans le choix de l'hébergement du stockage et du traitement de leurs données.**

Paralléliser vos opérations de test et de production en adoptant la “Livraison Continue”

» **Le principal objectif de l'utilisation de microservices, de conteneurs et de plateformes d'orchestration est de tendre vers l'automatisation complète du processus de développement.**

Pour reprendre la métaphore du travail portuaire, si la forme du conteneur est connue et comprise, si on arrive à en lire automatiquement le code barre, on peut imaginer la mise en place d'un “robot grue” récupérant de manière automatique le conteneur sur une digue pour le déposer sur un transbordeur.

Les conteneurs Kubernetes permettent d'oublier jusqu'à l'existence d'un centre de données



Source : Artefact

Les conteneurs et les maillages de services permettent d'associer de façon transparente les environnements on premise et cloud pour en obtenir la meilleure combinaison tout en capitalisant sur un même ensemble d'outils et de compétences

L'enjeu est d'arriver à développer des workflows automatisés qui (1) récupèrent la dernière version du modèle du data scientist (préalablement packagée sous forme de conteneur), (2) soumettent le conteneur à des tests de conformité pour s'assurer de son éligibilité à une mise en production, (3) lancent un processus continu d'industrialisation et de montée en version.

Sans aucune intervention humaine spécifique côté production, le code est comparé au dernier conteneur déposé sur le transbordeur, et est progressivement remplacé par celui-ci une fois les tests de conformité validés.

» **Le rôle du data engineer n'est plus de gérer les demandes sporadiques des data scientists, mais de se concentrer sur la gestion et la maintenance de la plateforme qui gère les demandes et orchestre les workflows automatisés de mise en production. Son activité est beaucoup plus linéaire, prévisible ; il a créé une "mise en production as-a-service" dont il est l'unique garant de la roadmap produit.**

Pour accélérer la transformation, développer une intelligence artificielle "as-a-service"

Utiliser la boîte à outils des solutions cloud

» **Les différents outils proposés par les fournisseurs de cloud démocratisent l'utilisation de l'intelligence artificielle et en font une initiative globale.**

Ces outils, intégrés au sein de plateformes cloud plus complètes, permettent la gestion de l'ensemble du cycle de vie d'un projet de machine learning, de la collecte de données jusqu'au déploiement du modèle en production.

Les solutions cloud proposent différents niveaux d'outils qui peuvent être évalués au regard de deux axes :

- complexité d'utilisation
- degré de personnalisation du modèle

En fonction des compétences présentes dans l'organisation et des spécificités des cas d'usages à traiter, il s'agit de choisir l'outil qui maximisera le temps d'exécution du projet ainsi que l'adéquation de la réponse au problème posé.

» **Si l'on prend l'exemple de la détection d'objet sur des images, la boîte à outils cloud permet de résoudre le problème de trois manières différentes :**

#Niveau1

Utilisez un modèle pré-entraîné sur étagère et appliquez-le directement sur vos données pour reconnaître les produits. Cette méthode est la moins complexe et la plus rapide car il n'y a pas besoin de construire ou d'entraîner un modèle. Si l'on prend l'exemple de l'API Cloud Video Intelligence de Google, elle permet aujourd'hui de reconnaître automatiquement 20 000 labels, et de stocker et d'indexer des vidéos en cinq fois moins de temps qu'autrefois.

Ces solutions répondent parfaitement aux besoins de cas d'usages standards, au niveau de complexité maîtrisé. Elles restent limitées pour la réalisation de cas d'usages plus complexes nécessitant le déploiement d'un modèle personnalisé.

Par exemple, les APIs pré-entraînées de reconnaissance d'image sont en capacité de distinguer un sac d'une marque X par rapport à celui d'une marque Y en s'appuyant sur la reconnaissance du logo, mais ne vont pas être en capacité de distinguer les différents modèles de sacs proposés par les deux marques.

#Niveau2

Entraînez un modèle sur vos propres données, afin que le modèle puisse apprendre et connaître précisément la gamme de vos produits. Ce niveau nécessite d'entraîner un modèle sur les données que vous avez labellisées au préalable. Cela prend plus de temps mais apporte des résultats plus pertinents car le modèle reconnaîtra les références exactes du produit.

#Niveau3

Créez votre propre modèle à partir de zéro. Dans ce niveau, vous n'entraînez pas seulement un modèle, vous le construisez dès le début. Il peut s'agir de l'ajustement de paramètres sur un modèle existant, ainsi que du développement d'une nouvelle architecture de réseaux neuronaux. Plus chronophage, ce niveau est généralement dédié au domaine de la recherche.

» Dans la mesure où ils répondent au besoin, l'utilisation de ces outils doit être maximisée :

- + **Pensés sur étagère, parfois pré-entraînés, leur utilisation réduit le time-to-market des projets IA et permet une meilleure gestion des ressources humaines**
- + **Conçus par les différents éditeurs de cloud pour être "by-design" scalable, ils garantissent l'accélération du déploiement de l'IA à l'échelle**

Utiliser des outils collaboratifs de création et visualisation de workflows data

» Le développement de projets d'intelligence artificielle est facilité par l'utilisation d'outils permettant aux équipes de :

- + **collaborer au sein d'un même environnement**
- + **suivre et visualiser intuitivement les différents workflows data et leurs états avancements**

Le déploiement sur des infrastructures cloud d'outils spécialisés comme la solution française Dataiku facilite le lancement d'initiatives IA au sein d'entreprises en début de transformation. C'est une plateforme d'Analytique et de Data Science centralisée qui accompagne les organisations dans leur transition vers l'intelligence artificielle d'entreprise. Elle réunit à la fois les experts data et les métiers, fournit des référentiels des meilleurs pratiques/modèles et offre des outils innovants pour la gestion et le déploiement du machine learning et de l'IA. En favorisant la collaboration, Dataiku offre une multitude de possibilités pour à la fois créer des modèles ayant un impact visible et rapide mais aussi construire les expertises data de demain.

» Alors que la mise en place d'une infrastructure scalable, modulaire et agile est un pré-requis pour le passage à l'échelle, les outils doivent être considérés comme des accélérateurs de la transformation. Conçus pour rendre accessible l'intelligence artificielle à des profils non experts, ils ont la vertu de rendre la transformation IA plus inclusive, accélérant sa diffusion à l'ensemble des salariés.

Lancer des IA labs pour accélérer et créer massivement de la valeur

Pourquoi mettre en place un IA lab ?

Qu'est ce qu'un IA lab ?

Un IA lab est une structure en charge de créer de la valeur via la réalisation de cas d'usages d'intelligence artificielle. Un IA lab regroupe :

- **des compétences pluridisciplinaires** : des profils à la fois business (product owner, business owner...) et techniques (data scientist, data engineer, data architect...) qui oeuvrent ensemble à la réalisation d'un ou plusieurs cas d'usages IA
- **des méthodologies** : Design Thinking, Agile, qui permettent de passer rapidement de l'idée au concept, du concept au prototypage, du prototypage à la mise en production et de la mise en production au déploiement à l'échelle
- **des technologies** : qui permettent de réaliser des solutions techniques performantes et scalables

Les grands enjeux

Les initiatives autour de l'intelligence artificielle sont souvent dispersées dans les organisations, la plupart restent à l'état de POC, ne délivrant in fine aucune valeur.

Ces différents POCs sont rarement cadrés par une stratégie globale, et de nombreux gisements de valeur restent inexploités.

L'IA lab porte 2 grands enjeux :

- Créer massivement de la valeur, en priorisant de manière fine et intelligente les cas d'usages, en accélérant leur industrialisation et leur passage à l'échelle
- Favoriser la transmission de savoir et la montée en compétences des talents de l'entreprise

Les 5 convictions d'Artefact pour la mise en place d'un IA lab

#1 Choisir les bons cas d'usages

Être équilibré dans les différents domaines business et dans les différents gisements de valeurs.

Aller chercher des victoires rapides sur des 'greenfields' (potentiel de valeur jusqu'alors non exploité), à la fois pour garantir une vraie valeur incrémentale et pour faciliter l'implémentation. Il est important que les premiers cas d'usages réussissent, plus qu'ils ne portent sur les plus gros potentiels de valeur.

Etre pragmatique, privilégier le facile à la fois sur le plan technique et sur le plan organisationnel (choisir les cas d'usages avec le moins de frictions avec les métiers).

#2 S'organiser en feature teams

Mettre en place des teams agiles, chacune en charge d'une grande problématique business, pilotée par un KPI. Exemple pour un retailer : une team agile performance magasin, une team agile performance supply chain...

Découper ces teams agiles en features teams, chacune en charge d'un sous-problème. Exemple : l'agile team supply chain, peut être décomposée en une feature team prédiction des ventes, une feature team sur l'automatisation du travail en entrepôt...

#3 Décomposer la complexité

Segmenter le sous-problème de chacune des features teams en sous-unités de base moins complexes.

Exemple : la feature team en charge de la prédiction des ventes, s'intéresse dans un premier temps à la prédiction des ventes des fruits et légumes et plus précisément à la prédiction des ventes de la tomate.

#4 Assurer la montée en compétences

Instaurer des rituels, pour favoriser la transmission des savoirs et la montée en compétences, par exemple :

- + **Tech tex** : organisation d'évènements hebdomadaires sous forme de retour d'expérience d'une équipe sur un point spécifique (un outil, une difficulté etc.)
- + **Pair programming** : mise en place de binômes au sein des features teams, qui travaillent ensemble sur le même code

Créer un 'lab academy', en charge de la mise en place et de l'animation d'un programme de formation au sein du lab.

#5 Rendre scalable

Etre en capacité d'augmenter rapidement la capacité des équipes du lab en créant de nouvelles teams agiles ou en ajoutant des features teams au sein des différentes teams agiles.

L'organisation de l'IA lab

Board IA Lab — Définit les grandes orientations, arbitre sur la roadmap et le pipe du lab



Sponsors

IA Lab Program Management — Définit et pilote la roadmap globale du lab



Directeur
du lab



Program
Manager



PMO

Agile teams

Team agile #1

Travaille sur des cas d'usage en lien avec une problématique business

Feature team #1.1

En charge d'un sous problème



Feature team #1.2

Feature team #1.n

Team agile #2

Travaille sur des cas d'usage en lien avec une problématique business

Feature team #2.1

En charge d'un sous problème



Feature team #2.2

Feature team #2.n

...

Team agile #N

Travaille sur des cas d'usage en lien avec une problématique business

Feature team #n.1

En charge d'un sous problème



Feature team #n.2

Feature team #n.n

Platform Team

Met à disposition des features teams des infrastructures et des data products standards¹

Est garante de la cohérence de l'écosystème global



Représentant
IT



Data
Architect



Product
Owner



Data
Engineer

Growth team

Challenge les features teams, mobilise les business owners, fait adopter les réalisations du lab, communique sur les victoires



Growth
manager



Change
manager

Partenaires :

Apportent méthodologie et technologie

Editeurs de solutions :

(Google, Amazon, Microsoft, IBM...)

— Cabinets de conseil

— Agences

¹voir les 4 partis pris d'Artefact pour optimiser ses infrastructures page 92, 93, 94

Source : Artefact

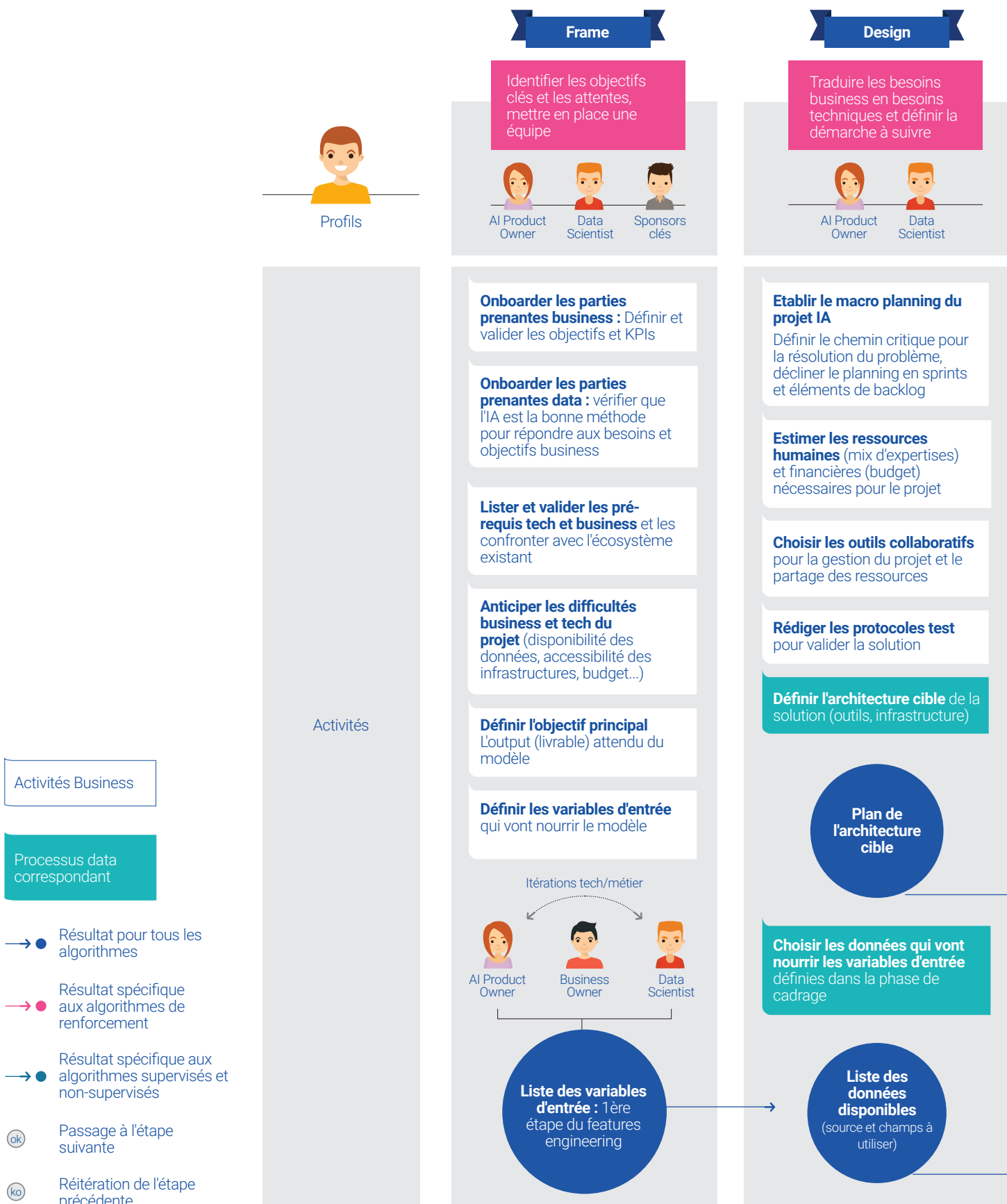
Le process de réalisation d'un cas d'usage

La réalisation d'un cas d'usage IA se décompose en 5 phases séquentielles mais itératives (il est toujours possible de revenir à plusieurs reprises sur une phase déjà engagée pour itérer sur le MVP). Notre méthodologie s'attache toutefois à contrôler à priori le nombre d'itérations pour garantir la bascule en production.

Les 5 étapes d'un projet d'intelligence artificielle

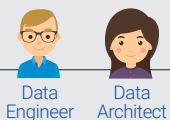
- **Encadrer** : identifier les objectifs et les attentes, définir précisément le périmètre du MVP
- **Designer** : définir le chemin critique de résolution du problème et le planning de delivery associé
- **Construire** : matérialiser le problème à résoudre (base d'apprentissage) et la solution pour y répondre (modèle)
- **Tester & Déployer** : valider que la solution répond au besoin business et déployer
- **Améliorer** : entraîner l'algorithme sur de nouvelles données, ne jamais cesser de tester pour s'améliorer

Ce schéma a vocation à fixer un cadre, lever le voile sur la complexité technique et organisationnelle inhérente à tout projet d'intelligence artificielle. Il offre au lecteur un fil d'Ariane à adapter aux spécificités propres à chaque organisation.



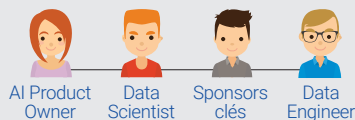
Build

Matérialiser le problème à résoudre et la solution pour y parvenir



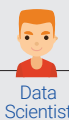
Test & Scale

S'assurer que la solution correspond aux besoins business et la déployer



Improve

Entraîner l'algorithme sur de nouvelles données (si des bases de données enrichies sont disponibles)



Préparer la base d'apprentissage (jeu de données)

Importer les données dans l'espace de travail

Construire la base d'apprentissage (utilisée pour entraîner le modèle)

Construire les variables cibles
Construire les variables explicatives (étapes 2 et 3 du features engineering)

Choisir la combinaison d'algorithmes qui servira à construire la solution (le modèle ML)

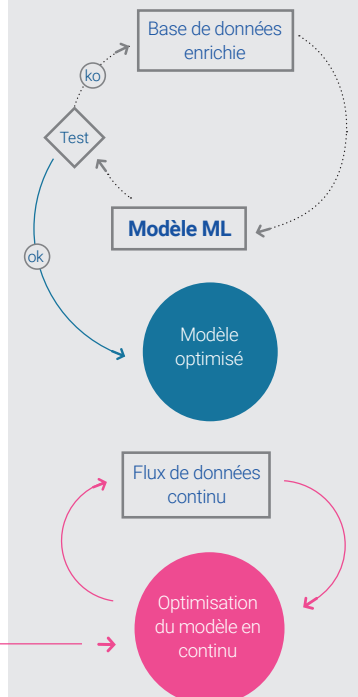
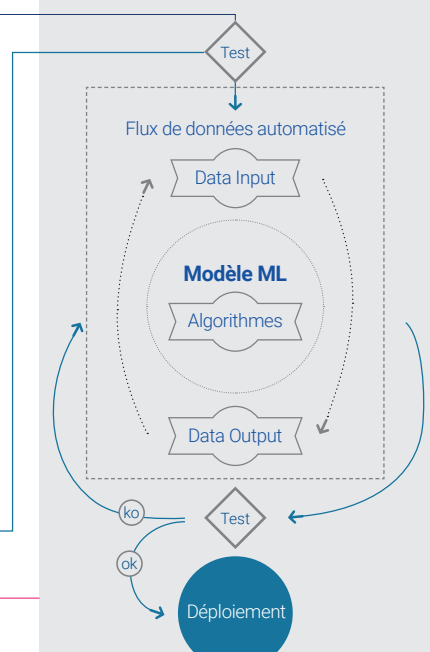
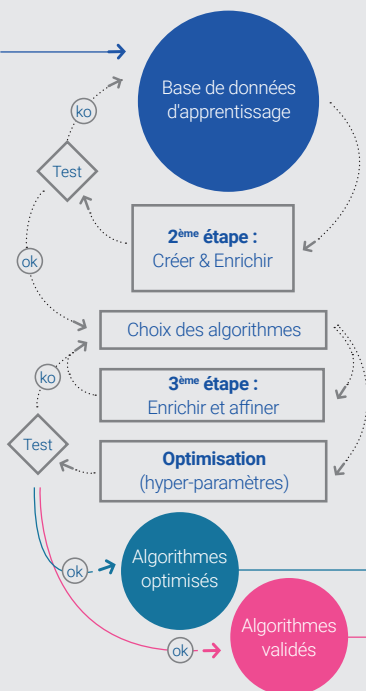
Optimiser la solution (le modèle ML)
- Features engineering*
- Affinage des hyper-paramètres

Intégrer le modèle ML dans son environnement de destination (ex : construction de flux de données)

Tests complets de la solution (ex: ramp-up tests*...)

S'assurer que la solution finale répond aux KPI et aux attentes business définies dans la phase de cadrage

Ré-entraîner le modèle



Source : Artefact

Développer une intelligence artificielle sécurisée et conforme

La protection des données, un enjeu clé

La protection des données, liée à la sécurité et à la confidentialité est souvent utilisée comme le critère discriminant pour arbitrer sur l'utilisation de plateformes cloud.

Pourtant, selon une étude Gartner publiée en 2016, d'ici 2020, les workflows hébergés sur le cloud public rencontreront 60% moins de failles de sécurité que les data centers traditionnels¹.

Sur le plan juridique, les principaux fournisseurs de cloud respecte les 8 grands principes du RGPD :

1. Licéité et loyauté
2. Transparence
3. Limitation des finalités
4. Minimisation des données
5. Exactitude
6. Limitation de la conservation
7. Intégrité et confidentialité
8. Responsabilité

Selon le RGPD, le data controller (entité qui collecte la donnée donc l'annonceur) est responsable du respect de ces principes. Toutefois, les principaux fournisseurs de cloud (data processor) proposent des approches pour se conformer aux exigences réglementaires, depuis une première phase d'évaluation jusqu'à la mise en conformité.

Même si d'un point de vue purement juridique, les données des entreprises pourraient être chargées librement sur le cloud, la principale barrière à l'adoption de ces solutions reste la réticence des annonceurs à confier leur patrimoine data à un tiers. Le principal point d'attention se situe autour de la sécurité, le cloud présentant des menaces propres au fonctionnement à la demande et partagé de ses infrastructures.

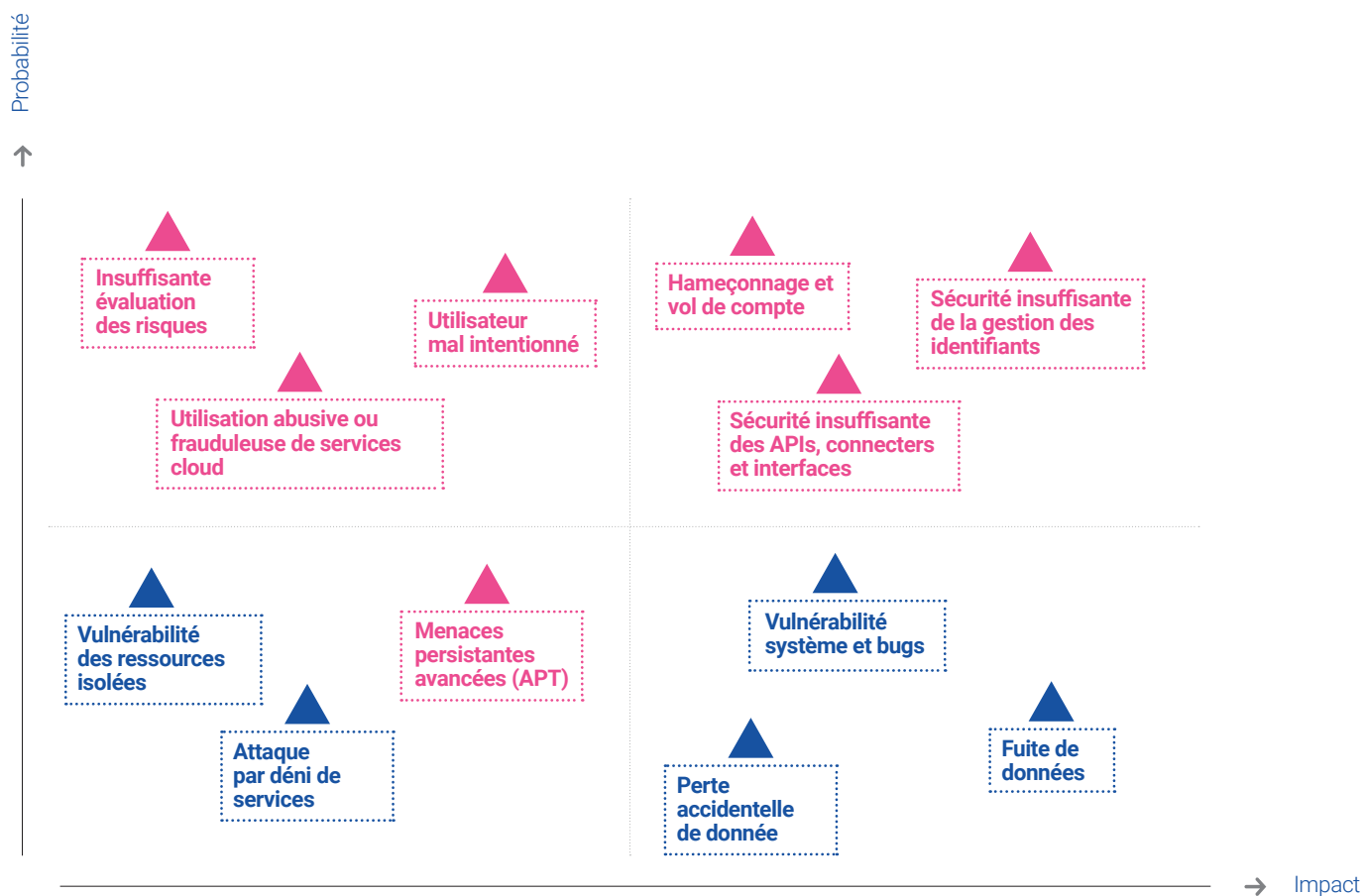
Pour répondre à cet enjeu sécuritaire, les principaux fournisseurs de cloud proposent des services et mesures garantissant à leurs clients une sécurité accrue.


¹Cloud are secure : are your using them securely ? Publié le 21 juillet 2016

Les 12 principales menaces associée à l'utilisation du cloud public

Aujourd'hui sur les questions de sécurité les menaces majeures sont principalement liées à l'humain.

Sans se vouloir exhaustive, la matrice ci-dessous présente les principales menaces propres au fonctionnement à la demande et partagé des infrastructures cloud.



 Menace avec probabilité forte liées à l'humain

Source : The Cloud Security Alliance (CSA) - "Trachercus twelve"

Les 5 chantiers à lancer pour sécuriser ses infrastructures et respecter les données de ses clients

Partant de cette cartographie, on distingue cinq principaux chantiers sécurité & privacy :

1. la confidentialité et souveraineté de la donnée,
2. la sécurité physique,
3. la sécurité des infrastructures,
4. la sensibilisation et gestion des utilisateurs,
5. la détection et gestion des incidents.

Ces différents piliers peuvent être traités par la voie de 3 leviers :

- **La contractualisation** : l'entreprise négocie avec le fournisseur de cloud pour avoir des garanties contractuelles autour de la sécurité des données et des infrastructures (exemple : les garanties autour de la sécurisation des datacenters du partenaire)
- **Les process internes et les outils** (cloud ou on premise) : l'entreprise s'outille et s'organise pour garantir une sécurité accrue (exemple : la mise en place de process d'anonymisation, la définition de rôles spécifiques, l'utilisation d'une brique d'authentification)
- **La confiance** : l'entreprise n'a pas d'autre levier que celui de la confiance envers son fournisseur de cloud

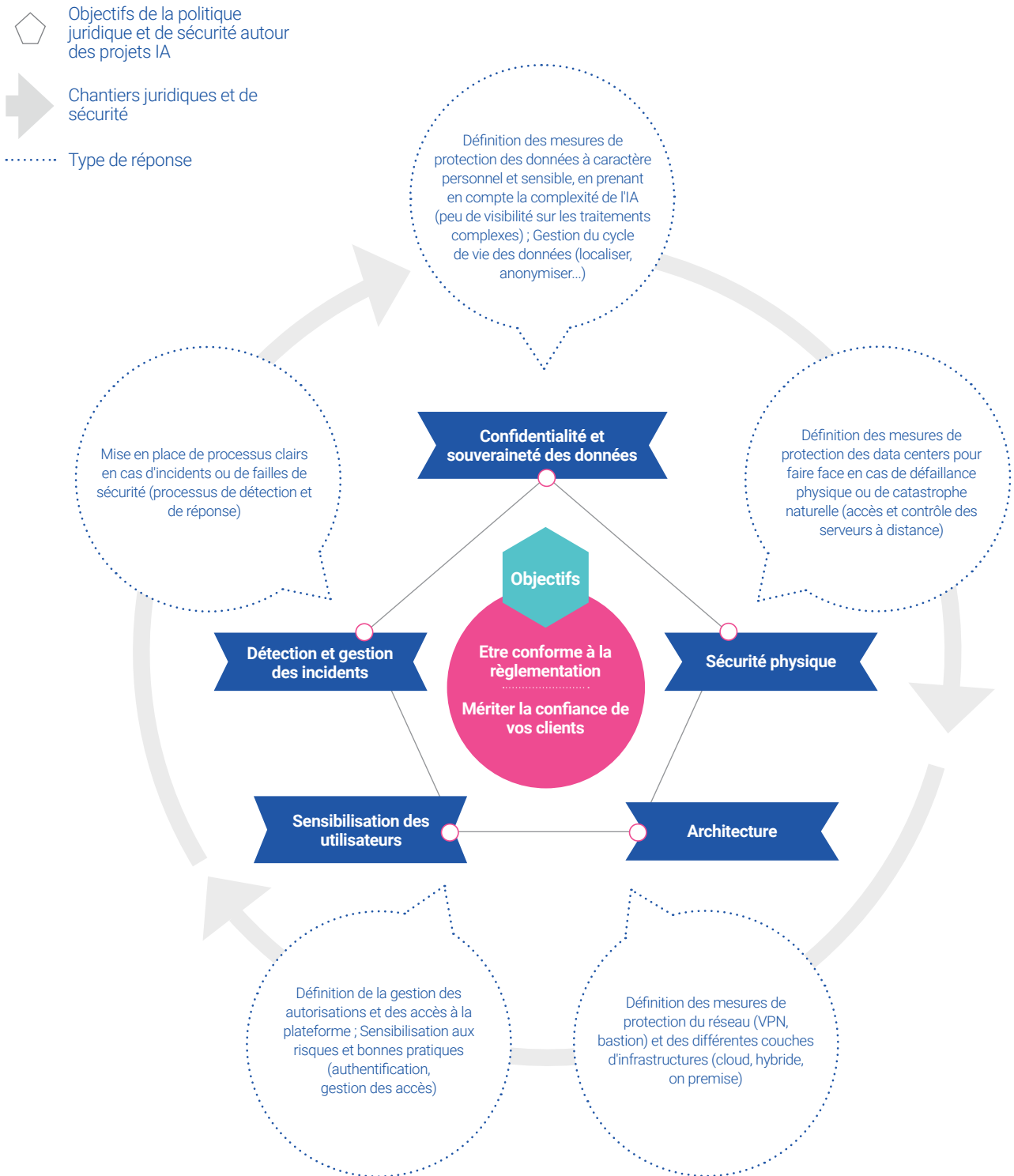
Pour chaque pilier il existe deux niveaux de protection :

- Les configurations proposées par les différents éditeurs de cloud, qui permettent de répondre en moyenne à 95% des exigences sécurité des grands groupes
- Les mesures spécifiques, mises en place par les directions sécurité internes qui couvrent les 5% restants

Pour Jean-Christophe Laissy, DSI de Véolia, le cloud public garantit un niveau de sécurité plus élevé que celui observé au niveau des data centers on premise.

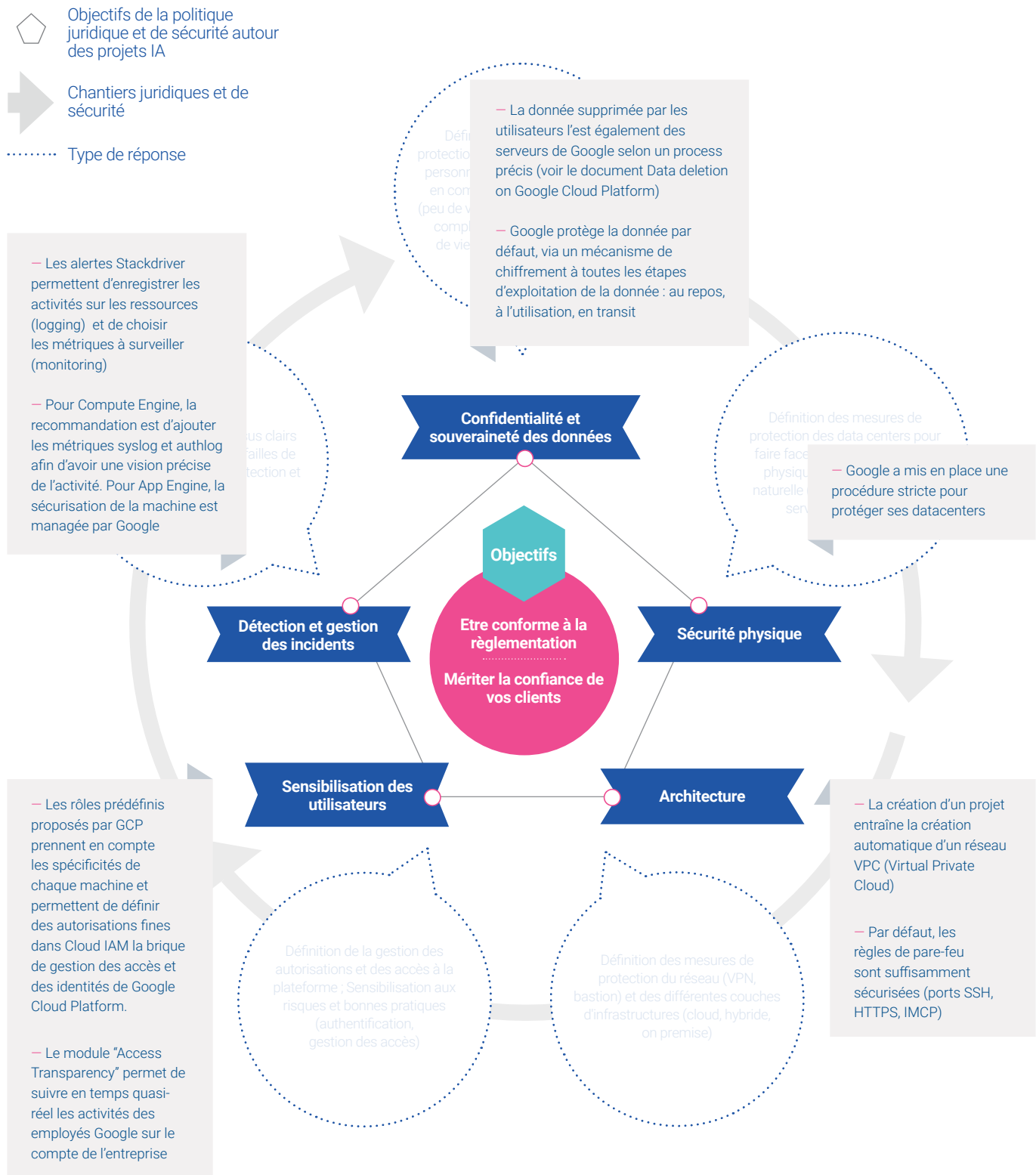
"J'espère aujourd'hui que plus personne ne pense que son data center interne est mieux protégé qu'un data center dans le cloud public contre les risques d'intrusion externe. On en est au même stade qu'avec les banques il y a une centaine d'années : plus personne ne pense que son argent est mieux protégé sous son matelas, qu'à la banque."

Mise en place du patrimoine du projet IA et des garanties de sécurité



Source : Artefact

Exemple de mesures proposées par Google pour garantir la sécurité des données sur Google Cloud Platform



Source : Artefact

02

Une fois la donnée libérée grâce à la mise en place d'un patrimoine data accessible et exploitable, et le développement d'une organisation tournée vers l'IA (chantiers culture & compétences), le passage de l'IA à l'échelle nécessite la mise en place d'enablers précis sur le plan technique et organisationnel.

Sur le plan technique, Artefact propose de bâtir l'IA sous forme de data products, construits via une philosophie de développement appelée microservices qui facilite l'industrialisation et le passage à l'échelle des projets d'intelligence artificielle. Les outils 'd'IA as a service' proposés par les différents fournisseurs de cloud, doivent quant à eux être perçus comme des accélérateurs de la transformation.

Sur le plan humain, l'enjeu est de mettre la donnée au coeur de l'entreprise, et se donner les capacités d'innover pour accélérer la transformation IA. Cela passe par la création de structures dédiées comme les IA labs, qui regroupent compétences, méthodologies et technologies, et ont pour objectif d'accélérer le lancement et l'industrialisation de cas d'usages IA à forte valeur ajoutée.

En bref.

Cependant, le principal frein que rencontrent aujourd'hui les entreprises dans la mise en place d'infrastructures hybrides ou full cloud sont les sujets liés à la privacy et à la sécurité des données. L'enjeu est de lancer en parallèle de la construction des enablers IA, nos 5 grands chantiers Sécurité et Privacy pour s'assurer de construire une Intelligence qui soit éthique, sécurisée et conforme :

1. la confidentialité et souveraineté de la donnée,
2. la sécurité physique,
3. la sécurité des infrastructures,
4. la sensibilisation et gestion des utilisateurs,
5. la détection et gestion des incidents.

Une fois les fondations de l'usine construites, il s'agit maintenant de s'attacher à optimiser sa chaîne de production et réduire les sources de gaspillage pour que l'IA devienne un objet de simplification et de productivité : il s'agit du "Lean".

03

Lean AI

Optimiser la chaîne de production de l'IA
pour maximiser la valeur

03

Selon une étude Artefact :

- 1/3 du temps, des efforts et des budgets des projets IA auraient pu être évités
- Plus de 3/4 des POCs ne sont pas industrialisés
- L'essentiel de la valeur est créée dans la seconde version des projets

Artefact propose une adaptation des méthodologies Lean Manufacturing, le Lean AI, pour optimiser la chaîne de production des projets d'intelligence artificielle.

Intro.

Le Lean AI repose sur l'élimination de l'ensemble des sources de gaspillage dans la chaîne de production. Cette méthodologie vise à augmenter la capacité des équipes à produire, en réduisant les coûts et le temps de réalisation du projet. Elle repositionne la compréhension des besoins du client au coeur de la démarche.

Dans un premier temps nous allons nous concentrer sur les 7 sources de gaspillage traditionnellement rencontrées, pour ensuite présenter en 3 axes la réponse apportée par le Lean AI.

Optimiser la chaîne de production IA

Les 7 “wastes” des projets d’Intelligence Artificielle

Artefact s’est inspiré des 7 “wastes” du lean manufacturing et les a adaptés au domaine de l’intelligence artificielle. Cette étude repose sur la réalisation de plus de 30 projets d’intelligence artificielle sur les trois dernières années.

Pour chaque “waste” identifié - Défauts, Talents, Extra Processing, Attente, Sur-production, Dépense d’énergie et Inventaire - les causes sont explicitées et éclairées grâce à des exemples concrets.

Extra Processing :

sur-délivrer par rapport aux attentes clients

L'IA pour la beauté du geste

Faire de l'IA pour faire de l'IA. Les équipes techniques s'attèlent à construire une solution technique efficace et innovante - car faisant appel aux dernières nouveautés technologiques - mais complexe. Les considérations business sont reléguées au second plan au profit du challenge technologique ; les spécifications techniques prennent le pas sur les spécifications métiers pas assez précises ; les fonctionnalités coeur génératrices de valeur ne sont pas identifiées. L'extra-processing aboutit souvent à la mise en place d'usines à gaz, de "boîtes noires", complexes à maintenir pour les équipes techniques et à comprendre et adopter pour les équipes métiers.

"Réinventer la roue"

Beaucoup de temps est consacré à la construction d'une IA personnalisée, au petit détail qui parfait la réalisation mais n'est pas attendu par le client. Le coeur du projet est délaissé au profit de l'accoutumance aux petits problèmes et petits défauts.

ex

— Un acteur de l'assurance souhaite créer un algorithme de prédiction du churn. Les algorithmes traditionnels, présents sur étagère, permettent de répondre au cahier des charges. Cependant, pour le challenge, l'équipe data souhaite concevoir une solution deep learning personnalisée. L'algorithme conçu est une "boîte noire", beaucoup plus long à produire et complexe à évangéliser.

— Un acteur des PGC souhaite mettre en place un algorithme d'attribution pour dé-siloter le pilotage des ventes média. Plutôt que de capitaliser sur des modèles d'attribution classique (last touch, linéaire, first touch etc.), le client souhaite mettre en place un modèle d'attribution data-driven, complexe d'appropriation pour les métiers.

Dépense d'Energie :

des efforts humains ou machine qui n'apportent pas de valeur incrémentale

Le manque d'intégration

Sur le plan technique, manque d'un écosystème IA articulé avec le SI legacy et les outils en place. Les différentes briques nécessaires à la construction du produit IA ne sont pas intégrées, générant des frictions dans le processus de développement.

Sur le plan humain, les compétences ne sont pas centralisées, la communication manque de fluidité entre les différentes parties prenantes. L'éparpillement des acteurs engendre perte de temps et quiproquo.

ex

— Un acteur du retail souhaite mettre en place un algorithme de recommandation client. La donnée nécessaire à la réalisation du projet (CRM, transactionnelle, navigationnelle...) est éparpillée dans plusieurs bases sans accès centralisé. Un chronophage travail de centralisation et de réconciliation des sources est nécessaire avant de lancer le projet.

— Pour bénéficier d'une puissance de calcul accrue, l'équipe data décide d'entraîner l'algorithme sur le cloud, et d'industrialiser le modèle côté on premise. Les écosystèmes techniques ne sont pas intégrés, ajoutant de nombreuses tâches au backlog et engendrant des efforts superflus côté développement.

Inventaire :

la création d'intelligence pas assez autonome

La culture du spécifique versus standardisé

Les modèles sont conçus via une chaîne de production très spécifique, répondant à un besoin particulier. Aucune capitalisation sur ce qui a été construit n'est possible. On recrée à partir de 0 à chaque nouveau besoin, ce qui ralentit la réalisation et mise en production de nouveaux modèles.

La dette IA

L'Intelligence construite n'est pas assez autonome, la maintenance est trop complexe. On a sur-produit de l'intelligence de manière non intelligente, tous les choix faits par le passé se révèlent être des fardeaux techniques. Les ressources sont mobilisées sur de la "vieille maintenance" plutôt que sur de la "nouvelle intelligence".

ex

— Un acteur des cosmétiques souhaite créer une solution de NLP pour faire de l'analyse sémantique sur les réseaux sociaux. Plutôt que d'utiliser une brique standard et d'y apposer une brique spécialisée pour l'analyse des réseaux, la solution n'est pas pensée pour servir d'autres cas d'usages. 3 mois plus tard, une autre équipe souhaite utiliser du NLP pour un autre type d'analyse, elle est forcée de repartir de 0.

— Un acteur de l'IoT souhaite piloter l'usage de ses objets connectés. La donnée en entrée est erratique et de très mauvaise qualité. Il externalise la mise en place d'une chaîne de production très spécifique, permettant de traiter la donnée en temps réel et d'anticiper au maximum les problèmes de qualité. Les équipes internes sont dans l'incapacité de reprendre et maintenir cette solution trop spécifique et complexe.

Sur-Production :

l'IA arrive trop tôt par rapport à la maturité de l'entreprise, ou de manière non structurée, non réfléchie

La culture du POC non industrialisable

Le manque d'une stratégie globale et de synergies entre les différents acteurs de l'organisation, fait naître de nombreuses initiatives disparates (POCs) en dehors des principales poches de valeur de l'entreprise.

Ces diverses initiatives arrivent souvent en anticipation des attentes clients, et plusieurs répondent parfois à une seule et même problématique.

ex

— Une entreprise du secteur des PCG décide de lancer un laboratoire autour de l'intelligence artificielle. La première activité du lab est l'organisation d'ateliers d'idéation pour identifier des cas d'usages éligibles à une entrée en lab. Les cas d'usages choisis, sont principalement axés promotion et marketing édulant d'autres poches de valeur comme l'optimisation de la supply chain, la réduction de la consommation énergétique, la réduction des coûts en magasin etc.

Attente :

temps inoccupé, dans l'attente d'une livraison ou d'une décision managériale

Le manque de sponsoring C-level

L'intelligence artificielle n'est pas considérée comme un levier stratégique par les dirigeants de l'entreprise. Le temps passé par le management à convaincre que l'intelligence artificielle peut créer un avantage concurrentiel, laisse peu de temps pour l'action.

Le faible sponsoring des dirigeants entraîne un manque de mobilisation de la part de l'ensemble des équipes, déséquilibrant les processus en place. Ceux-ci sont peu qualitatifs (problèmes de qualité, présence de défauts, de bugs techniques) et peu fluides avec un temps conséquent perdu entre deux étapes successives.

ex

— Un acteur du voyage souhaite développer une application à la voix sur Google Home et Alexa. Trois mois sont consacrés à l'identification et à la mobilisation de ressources techniques pour rejoindre l'équipe coeur. Le post mortem à trois mois témoigne d'un besoin fort de montée en compétence des équipes. L'entreprise décide de faire appel à des ressources externes, qui sont mises en attente un mois le temps d'ouvrir les accès aux bases de données et à l'environnement technique. Une fois le projet prêt à être lancé, il est mis en pause pendant plusieurs semaines dans l'attente de la signature d'un NDA par le CEO.

Défauts :

le produit IA ne répond pas aux besoins (dévie des attentes du client final)

Un focus sur la vitesse d'exécution versus la qualité et la cohérence

- Mauvaise compréhension du besoin métiers, spécifications pas assez précises
- Processus non standardisés, mauvaise détection des bugs lors de la phase de production
- Fonctionnalités défectueuses car machines et outils non maintenus
- Récompense des équipes sur la quantité et la rapidité d'exécution versus la qualité

ex

— Une entreprise du secteur des PGC souhaite remplacer son dispositif de diagnostic des pannes frigorifiques - aujourd'hui construit sur la base d'un arbre de décision déterministe - par un algorithme de machine learning. Le résultat obtenu n'apporte pas de valeur incrémentale versus le système en place ; plutôt que de laisser l'algorithme définir la solution la plus pertinente pour résoudre le problème, l'équipe data force l'IA à suivre une logique humaine, reproduisant en l'améliorant à la marge le dispositif existant.

Talents :

échec à capitaliser sur les compétences et les connaissances internes

Manque d'une culture inclusive, qui positionne l'humain au coeur

Pour réussir sa transformation IA, l'entreprise doit s'attacher à diffuser une culture qui reconnaît les forces et la contribution de chacun des salariés, qui mobilise, et fait de la transformation IA une transformation inclusive. Aujourd'hui, trop peu de temps et de ressources sont consacrés à la montée en compétence, à la formation et au développement des salariés. Le transfert de connaissance au sein de l'entreprise, pourtant clé, est souvent difficile à mettre en place.

ex

3 ans après le lancement d'un grand plan de transformation IA, une entreprise du secteur bancaire consulte ses employés pour en dresser le bilan intermédiaire sur le plan humain. 3 principaux axes remontent de la consultation :

- Résolution de problèmes réalisée en chambre par les experts, ignorant la contribution des autres membres de l'équipes,
- Des idées d'amélioration imposées par le management ou par les experts au reste de l'équipe plutôt que conçues avec elles,
- Des produits conçus « au rabais » par manque de compétences, de connaissances ou par manque de mobilisation des équipes prises par le run ou pas assez « incentivées » sur la livraison du projet.

La réponse du Lean AI aux “7 wastes” des projets d’Intelligence Artificielle

Artefact propose une adaptation des méthodologies Lean Manufacturing pour remédier au 7 sources de gaspillage traditionnellement rencontrées dans les projets d’intelligence artificielle.

» La réponse du Lean IA se structure en deux axes :

- + **L’optimisation de la chaîne de production via le développement de data products**
- + **Un emphase forte sur l’adoption et la distribution des data products**

Si ces guidelines ont vocation à éviter les écueils et à optimiser la transformation par la réduction des pertes et l’adoption du produit final, elles ne sont pas figées et ont vocation à être adaptées aux contraintes et caractéristiques de chaque organisation.

Les résultats recherchés sont l’optimisation du time-to-market, la livraison continue, la fluidification de la collaboration et l’accélération des prises de décisions au service de la valeur.

L’optimisation de la chaîne de production via le développement de data products

La standardisation : une contrainte culturelle, source des principales avancées technologiques

Dans son ouvrage [The Design of everyday things](#), Don Norman observe que les avancées technologiques proviennent des technologies elles-mêmes ou de la standardisation. Il fait le parallèle avec l’histoire de l’automobile ; les premières voitures étaient difficiles à opérer, exigeant de la force et des compétences spécifiques. Certaines difficultés furent traitées grâce à la standardisation : l’invention du moteur, du starter et de l’embrayage par exemple.

La standardisation peut être assimilée à une contrainte culturelle. Une fois la phase d’apprentissage terminée, le jeune conducteur est confiant dans sa capacité à conduire n’importe quelle voiture partout dans le monde. La standardisation permet ainsi des avancées technologiques et fonctionnelles majeures.

Les data products : des Lego intelligents

» **Au coeur du Lean AI, la culture de la capitalisation et de la standardisation. Le Lean AI s’attache à construire des briques technologiques ouvertes sur l’écosystème, appelées data products, sur lesquelles l’entreprise puisse se reposer pour construire une stratégie IA cohérente et pérenne.**

L’objectif est de mettre à disposition des équipes data une bibliothèque de briques technologiques standards, pour que les projets aux mêmes caractéristiques techniques soient traités via la même approche et s’intègre de manière cohérente au reste de l’écosystème.

Pour être réutilisable, l’IA doit être “packagée”, pensée comme une structure en Lego, composée de plusieurs briques qui par combinatoire recomposent un ensemble plus complexe. Le modèle d’intelligence artificielle repose sur un socle de dénominateurs communs, des data products génériques, indispensables au bon fonctionnement de l’IA mais non spécifique à la problématique traitée, et sur des data products “spécialisés” faisant la spécificité du produit IA en question.

Uber s’est lancé dans la construction de ses dénominateurs communs en créant une plateforme de datascience appelée Michelangelo dont l’objectif est d’adresser les problèmes de production en standardisant les workflows et les outils pour l’ensemble des équipes data. La plateforme permet de facilement construire et opérer des systèmes de machine learning à l’échelle, en mettant à la disposition des datascientists des morceaux de code pré-packagés et la documentation associée pour favoriser le partage et la diffusion des connaissances.

Le modèle d'Intelligence Artificielle repose sur un socle de dénominateurs communs, des data products génériques, et des data products "spécialisés"

Si l'on prend l'exemple d'un retailer, les référentiels clients ou magasins sont des dénominateurs communs, des data products génériques qui peuvent être consommés par plusieurs data products spécialisés comme des algorithmes de recommandation produit ou des algorithmes de prédiction de l'écoulement des stocks.

La philosophie technique permettant la mise en pratique de ces data products sont les microservices (détaillés en partie 2).

Cette approche a de multiples bénéfices :

- Elle **favorise la montée en compétence** des équipes en standardisant les approches et en capitalisant sur les compétences internes
- Elle **réduit les défauts et augmente la qualité** du produit final en érigeant des standards de développement supportés par des contrôles qualité
- Elle **améliore le time-to-market** des projets grâce à la centralisation des connaissances et la forte capitalisation sur les expériences passées

» Elle permet aux équipes de se concentrer sur l'essentiel : les 20% spécifiques qui vont permettre de décupler la valeur générée par le produit final.

Une emphase forte sur le dernier kilomètre

La méthodologie Lean IA met l'emphase sur le dernier kilomètre de la chaîne de production : un data product n'est abouti que s'il est distribué et consommé largement par les utilisateurs finaux.

La distribution des data products

Le travail des équipes data ne s'arrête pas à la construction du modèle mais à l'observation d'un incrément ou décrétement sur le KPI que l'on cherche à optimiser. Cela passe par mise à disposition de l'algorithme à l'ensemble des utilisateurs finaux.

Si l'on prend l'exemple de Google, une fois un nouveau modèle conçu, il est packagé sous forme de produit pour que les utilisateurs de la suite Google Cloud Platform puissent consommer le modèle avec le moins de frictions possibles. Cela passe par la mise en place des APIs machine learning créées pour développer l'usage des produits d'IA de Google. L'objectif des data products est de bénéficier à une audience bien plus large que celle pour laquelle le produit a été initialement conçu.

Un modèle de machine learning permettant de prédire la vitesse d'écoulement des stocks pour un EAN (European Article Number) donné, peut être utilisé largement chez un retailer s'il alimente par API (data product générique ou standard) un dashboard mis à disposition de l'ensemble de l'entreprise.

Ce dashboard peut être utilisé par les directeurs de magasins pour suivre l'écoulement de leurs stocks, par l'équipe marketing pour cibler ses campagnes sur les produits qui s'écoulent plus lentement, par l'équipe promotion pour définir sa stratégie promotionnelle, par l'équipe supply pour optimiser la chaîne d'approvisionnement etc.

L'IA Analytics

Une fois le data product diffusé largement aux différents acteurs de l'entreprise, il est nécessaire d'en mesurer l'adoption et d'en piloter la performance.

» Artefact préconise la mise en place d'un outil d'IA Analytics, à la main du Product Owner, pour favoriser l'amélioration continue du modèle.

Exemple de KPIs d'IA Analytics :

- **Des KPIs d'usage** (ex. nombre d'acteurs qui consomment le modèle)
- **Des KPIs techniques** (ex. performance du modèle (% d'erreur), nombre de défaillances sur la chaîne de production)
- **Des KPIs business** (ex. économies générées, incrément de chiffre d'affaires)

L'IA Product Owner ethnographe

Pour de l'analyse plus qualitative, les sessions d'immersion avec les équipes métiers sont préférées à l'envoi de sondages utilisateurs. L'IA product Owner agit à la manière d'un ethnographe, et observe comment l'outil créé est utilisé dans son environnement de travail.

Cette méthodologie s'inspire des bonnes pratiques observées pour le développement de produits digitaux. Pierre Fournier, Head of Product chez ManoMano nous partage sa méthodologie de feedback loop : les product owners organisent des "vendredis popcorn" au cours desquels ils visionnent des enregistrements vidéos de parcours utilisateurs sur leur site web. Le premier vendredi du mois est dédié au visionnage de 20 sessions sur le parcours de paiement, et le dernier du mois au visionnage de 20 sessions sur l'arrivée sur le site. Ces cérémonies rituelles permettent aux Product Owners de s'identifier aux utilisateurs et de percevoir les éventuelles difficultés ressenties en "immersion", ce qui leur permet de prioriser de manière plus juste les ajustements produits.

Le plan d'action du CAIO

La roadmap



LAUNCH AI : Lancer des initiatives, définir sa stratégie

Apprendre et expérimenter

- Monter en compétence
- Lancer des POCs

Définir les ambitions stratégiques

- Identifier et anticiper les poches de valeur internes (assets et synergies) et externes (marché, concurrence, ou différenciation)
- Structurer les investissements (opportunités technologiques...)
- Construire une stratégie IA (make, buy, both)

Construire son plan de transformation IA

- Communiquer la vision
- Construire une réponse organisationnelle
- Définir le plan de transformation
- Piloter la transformation (OMTM "one metrics that matters")



SCALE AI : Bâtir de solides fondations pour faire de l'impact à l'échelle

Libérer la donnée

- Construire un patrimoine data accessible et exploitable
- Acculturer et recruter

Mettre en place les bons enablers

- Mettre en place des infrastructures et des outils 'AI friendly'
- Lancer des IA labs

Développer une IA sécurisée et conforme

- Initier 5 grandes chantiers : confidentialité et souveraineté, sécurité physique, sécurité des infrastructures, sensibilisation et gestion des utilisateurs, détection et gestion des incidents

Industrialiser les POCs

Passer à l'échelle les POCs à valeur priorités



LEAN AI : Optimiser la chaîne de production de l'IA pour maximiser la valeur

Monitorer l'efficacité de l'IA

- Back-loop : revenir à la préparation, poursuivre l'innovation, ajuster les ambitions et le plan de transformation, puis adapter les fondations

Hyper-optimiser

Défaut

Attente

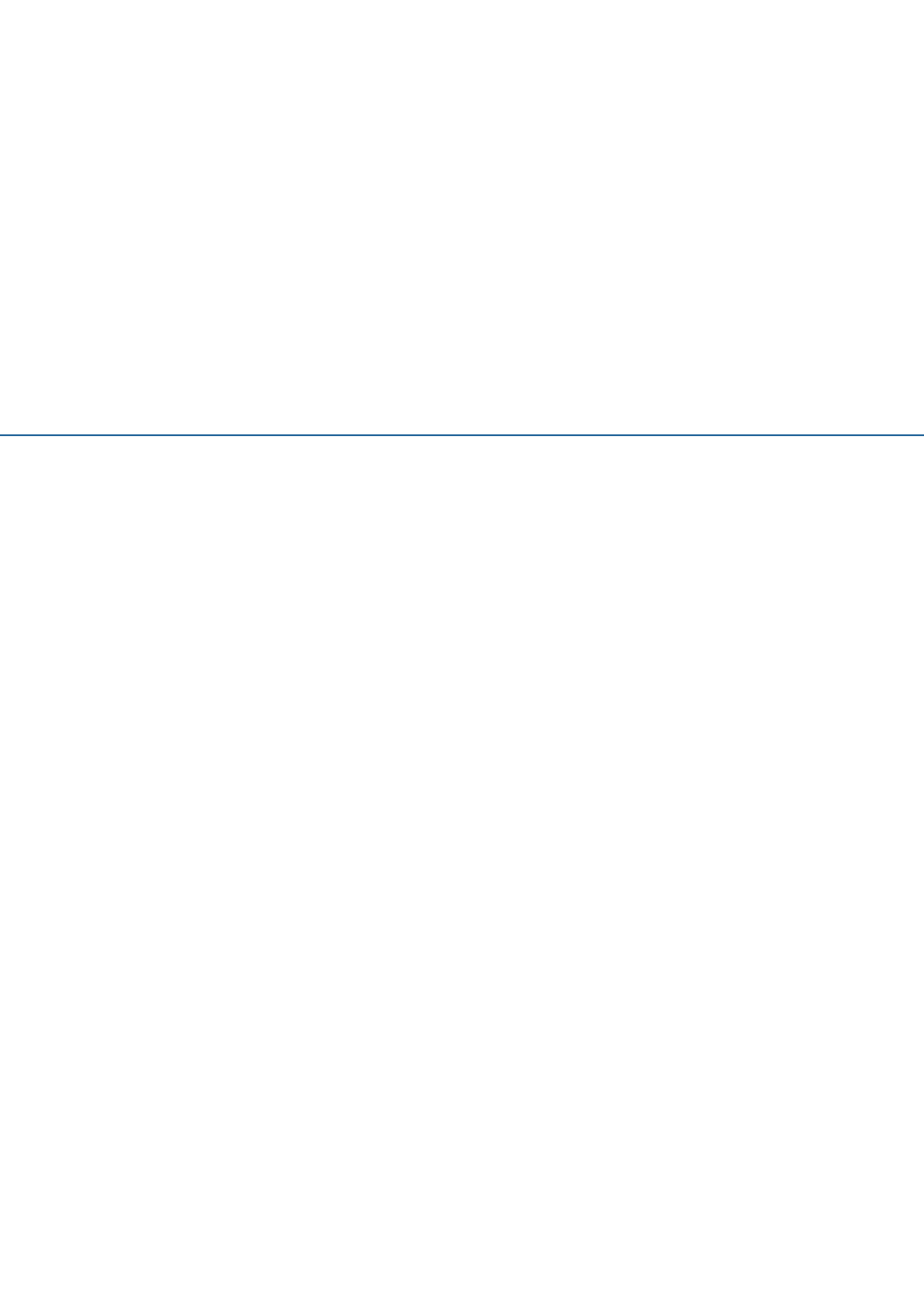
Talents

Inventaire

Surproduction

Dépenses d'énergie

Extra processing



Conclusion,

L'intelligence artificielle (IA) nous fait entrer dans une nouvelle révolution... une révolution à la fois sociale, culturelle, (géo)politique... et économique.

Avant l'explosion des usages digitaux, les marques s'imposaient par leur réseau physique, puis par l'audience de leur site internet, plus tard de leur application / site mobile. Les marques de demain s'imposeront par leur IA comme leur nouvelle formule Coca-Cola, instrument indispensable de leur compétitivité. C'est l'IA qui créera les usages, c'est l'IA qui gouvernera les flux logistiques, c'est encore elle qui sous-tendra le lancement de nouveaux services et produits.

Mais cette ère de l'IA pour les organisations sera surtout celle "des IAs". Des centaines, des milliers d'IAs devront être créées (car l'IA demande un véritable travail de création) et développées pour répondre au besoin de propagation de l'intelligence artificielle. L'ampleur du phénomène demande des outils d'accélération, qui permettent de dépasser les difficultés connues : rigidités des organisations, pénurie de compétences, qualité des données insuffisantes, vétusté des systèmes informatiques, etc.

Le cloud est un de ces outils d'accélération, déjà aujourd'hui considéré comme indispensable par les organisations qui font de l'intelligence artificielle une priorité¹. Ce plébiscite massif du cloud par les champions de l'intelligence artificielle dépasse le débat sur le dé plafonnement du hardware. Le cloud est le véhicule d'une "culture IA" qui mutualise les efforts d'apprentissage, les efforts de modélisation et les efforts d'intégration.

Mais le cloud ne sera lui-même qu'une étape. L'étape suivante, qui se présente déjà comme inévitable, est celle de l'ouverture des intelligences artificielles les unes aux autres au sein de réseaux d'IAs, gouvernementales et industrielles, nationales et transnationales. Cette "Open AI" qui succède au "Lean AI" annonce un nouveau modèle économique, celui de la coopération où les grands gagnants de demain seront ceux qui tireront le meilleur parti de la libération des IAs.

Et alors de voir émerger, dans les 3 à 4 années qui viennent, les premières blockchains IA qui nous feront basculer d'une économie du software, à une économie du netware. Les entreprises qui souhaitent devenir les nouveaux leaders de cette économie, doivent se lancer dans des bouleversements profonds de leur fonctionnement, et ces bouleversements qui commencent maintenant.

¹ Etude d'Artefact sur l'adoption du cloud en 2018 réalisée auprès de 100 entreprises dont les sièges sociaux sont situés en Europe, Afrique et Asie

Glossaire,

A

Algorithme :

Un algorithme est un format de réponse mathématique à un problème : c'est un ensemble de règles à suivre pour résoudre des opérations en un nombre fini d'étapes. Il existe une très grande variété d'algorithmes d'IA, pouvant être utilisés dans le cadre de plusieurs logiques d'apprentissage automatisés. Par exemple : réseaux de neurones, arbre décisionnel, régression logistique...

API :

Une API (Application Programming Interface) est une interface de programmation applicative. Plus concrètement, il s'agit d'une brique de code développée par un tiers qui lui permet de mettre à disposition de manière normalisée, des fonctionnalités, sans que les utilisateurs aient à comprendre le code sous-jacent. Par exemple, Google a mis à disposition son API Google Maps qui permet à des applications et des sites web d'utiliser la carte google.

Apprentissage non supervisé :

L'apprentissage non supervisé est une méthodologie d'apprentissage du machine learning qui permet d'identifier des tendances ou de regrouper des ensembles qui présentent des similarités. L'apprentissage non supervisé, est le mode principal d'apprentissage des humains et des animaux. C'est en observant le monde et en s'adaptant que l'on apprend, par nous-même. Son objectif est l'exploration à partir de données non classées.

Apprentissage par renforcement :

L'apprentissage par renforcement est une méthodologie d'apprentissage du machine learning, qui se différencie par son apprentissage continu : le modèle n'a jamais fini d'apprendre et se met à jour automatiquement pour améliorer sa "performance" tout au long de son usage. Le modèle apprend lui-même de ses actions. Les données qui servent à son apprentissage sont présentées une à une de façon successive en fonction de leur disponibilité.

Apprentissage semi supervisé :

A la frontière entre les logiques supervisées et non supervisées, on utilise une approche dite semi-supervisée qui permet de s'affranchir en partie des étapes de labellisation qui peuvent être fastidieuses sur des jeux de données importants.

Apprentissage supervisé :

L'apprentissage supervisé est une méthodologie d'apprentissage du machine learning, qui permet de prédire la valeur d'une variable cible à partir d'un entraînement sur une base de donnée existante labellisée. Pour faire cet apprentissage le data scientist s'appuie sur des algorithmes comme la régression linéaire, la régression logistique, l'arbre de décision, les réseaux de neurones etc. Son véritable intérêt réside dans sa capacité de généralisation, soit sa capacité à classer correctement des images d'objets qu'elle n'a jamais vues durant la phase l'apprentissage.

Approche déterministe :

L'approche déterministe est une réalité dans laquelle se succèdent des phénomènes dont les paramètres sont entièrement connus et maîtrisés.

Approche probabiliste :

L'approche probabiliste modélise une représentation statistique de la réalité : un environnement où des événements ont une probabilité d'occurrence.

Architecture :

Une architecture est la structure générale d'un système informatique. C'est l'organisation des différents éléments qui le compose, et leurs interactions. Ce sont les data engineers qui pensent, conçoivent l'architecture finale d'un projet et anticipent l'ensemble des contraintes liées au déploiement du projet IA. Ils doivent faire partie de l'équipe coeur dès le début du projet.

C

Calcul distribué :

Le calcul distribué est un calcul ou un traitement informatique qui est réparti sur plusieurs microprocesseurs. Le calcul distribué a révolutionné le monde du calcul informatique et le développement de l'intelligence artificielle, en fournissant des ressources de calculs quasiment infinies pour un coût modique.

Cloud computing :

Le cloud computing est une infrastructure informatique dans laquelle la puissance de calcul et le stockage sont gérés par des serveurs distants auxquels les usagers se connectent via un réseau sécurisé. Il peut être privé ou public et hybridé ou non avec des infrastructures on-premise.

Cloud hybride :

Le cloud est dit "hybride" ou "meilleurs des deux mondes" lorsqu'il combine cloud public et cloud privé (ou infrastructure locale "on premise").

Cloud privé :

Le cloud est dit "privé" lorsque les ressources cloud (serveurs, stockage etc) sont utilisées de façon exclusive par une organisation, via un réseau privé ou internet. Il peut être situé physiquement dans l'organisation (cloud interne) ou hébergé par un fournisseur tiers. Si toute l'infrastructure n'appartient pas à l'organisation, la maintenance des services et de l'infrastructure est de sa responsabilité. Par rapport au cloud public, le cloud privé est personnalisable : il s'adapte aux besoins.

Cloud public :

Le cloud est dit "public" lorsque les ressources cloud (serveurs, stockage etc) sont mises à disposition par un fournisseur tiers et livrées via Internet. Toute l'infrastructure reste la propriété du fournisseur de solution cloud. Les ressources sont mutualisées entre plusieurs organisations tout en assurant l'étanchéité entre chacune d'elles.

Conteneur :

Un conteneur est une enveloppe virtuelle qui permet de packager une application avec tous les éléments dont elle a besoin pour fonctionner : fichiers source, bibliothèques, outils et fichiers. Ils sont packagés en un ensemble cohérent et prêt à être déployé sur un serveur et son système d'exploitation.

Couche d'abstraction :

La notion de couche d'abstraction permet de décrire les systèmes informatiques comme une suite de couches se superposant. Chaque couche apporte de nouvelles fonctions de plus en plus élaborées et repose sur les fonctions plus élémentaires assurées par les couches sous-jacentes.

En cloud computing, les couches d'abstractions correspondent à des tâches de plus en plus abstraites déléguées sur les infrastructures tierces (cf. schéma Les 4 unités d'échelle du cloud computing p.63).

CPUs :

Les CPUs (Central Processing Units), ou processeurs, sont des systèmes permettant l'exécution des instructions d'un ordinateur.

D

Datalake :

Un data lake est une plateforme technologique utilisée dans les projets Big Data. Elle permet de centraliser l'intégralité des données brutes (structurées ou non) et hétérogènes (image, audio, texte) générées par une entreprise au sein d'une seule et même plateforme. Cette dernière permet non seulement de stocker et croiser les données, mais également de les manipuler et de les analyser.

Deep Learning :

Le deep learning ou apprentissage profond, est un algorithme spécifique du machine learning. On parle de réseaux de neurones profonds, à partir d'un certain nombre de "couches" de profondeur du réseau. En faisant un parallèle avec le cerveau humain, plus le nombre et la densité de neurones impliqués entre une information en entrée et une information en sortie est important, plus le réseau est considéré comme profond.

Design thinking :

Inspirée de la démarche suivie par les designers, le design thinking est un ensemble de méthodes "d'intelligence collective" permettant de résoudre des problèmes en recentrant les besoins utilisateurs au cœur de la réflexion et du développement de la solution (produit, service, expérience, forme de management...).

F

FaaS (Function as a Service, en anglais) :

Le FaaS est un modèle de service de cloud computing. L'infrastructure physique, le système d'exploitation et l'environnement applicatif sont fournis par la solution cloud. Les applications sont proposées sous forme de services clés en mains auxquels les utilisateurs se connectent via des logiciels dédiés ou un navigateur Internet. AWS Lambda, Google Cloud Fonction ou encore Microsoft Azure Functions en sont des exemples. (cf. schéma Les 4 unités d'échelle du cloud computing p.63).

Features Engineering :

Le features engineering est le processus qui permet de transformer de la donnée brute en données "comestibles" pour l'algorithme : des variables dites explicatives ou features. Il se compose de 3 étapes : 1, identification des données nécessaires à la matérialisation et résolution du problème posé ; 2, préparation et mise en qualité de la donnée choisie ; 3, construction des variables explicatives par combinaison des données choisies (à l'aide d'opérations mathématiques simples).

Framework de calcul :

En pratique les calculs et traitements informatiques sont découpés en une multitude de calculs et traitements secondaires, appelés framework de calculs.

Par exemple, Hadoop est un framework de calcul qui permet d'effectuer des traitements Big Data. Spark et MapReduce sont également des framework de calcul et reposent tous deux sur le principe du calcul distribué. Pour Spark, le stockage de la donnée se fait dans la mémoire vive et non plus sur le disque dur comme pour MapReduce, ce qui permet un traitement plus rapide.

G

Google Cloud Vision :

Google Cloud Vision est une brique IA du catalogue IA de GCP, disponible sous forme d'API, qui est utilisée spécifiquement pour la reconnaissance d'image.

GPUs :

Les GPUs sont de puissants processeurs spécialisés, initialement conçus pour le rendu graphique des jeux vidéo, qui se sont avérés être très performants pour les calculs des réseaux neuronaux. Ces GPUs sont principalement utilisés pour la phase d'apprentissage des réseaux de neurones.

IA faible :

Il existe plusieurs types d'IA. Aujourd'hui, l'IA dites "faible" est celle sur laquelle la majorité des recherches sont concentrées, et c'est celle que nous retrouvons dans la plupart de nos objets du quotidien. L'IA faible est un ensemble de programmes développés pour des applications spécifiques : le jeu de go, les échecs... On parle d'IA faible, car elle effectue ce que l'homme lui a appris à faire dans un domaine déterminé. Elle est donc mono-tâche, et automatise une tâche souvent répétitive pour l'homme.

IA forte :

Il existe plusieurs types d'IA. Aujourd'hui, nous sommes encore loin du stade d'une IA forte. Cette dernière serait capable de raisonner de façon abstraite, de déterminer ses propres objectifs, de résoudre plusieurs problèmes simultanément et de ressentir et exprimer des émotions. Ayant acquis le sens commun, et ayant conscience d'elle-même, elle serait capable de se reprogrammer et de ré-écrire son propre "algorithme cognitif".

IaaS (Infrastructure as a Service, en anglais)

L'IaaS est un modèle de service de cloud computing. L'infrastructure physique est fournie par la solution cloud. Le système d'exploitation et les applications sont installés par les clients sur des serveurs auxquels ils se connectent pour travailler comme s'il s'agissait d'un ordinateur classique. Amazon Web Services ou encore Google Cloud Engine en sont des exemples. (cf. schéma Les 4 unités d'échelle du cloud computing p.63).

Infrastructure :

L'infrastructure est l'environnement capacitaire pour l'élaboration d'une solution. L'infrastructure est souvent pensée "multiples", afin d'assurer l'agilité et la flexibilité de la solution finale. Il existe plusieurs typologies d'infrastructures : l'infrastructure d'exploration permet de prototyper le produit data lorsque l'infrastructure de production permet de faire passer le modèle du simple prototype à un produit consommé par des services ou applications tierces. On distingue également les infrastructures "on premise" physiques, des infrastructures cloud, qui sont des espaces de stockage et de calculs tiers.

Infrastructure on premise :

Le terme "on premise" (sur site) fait référence aux infrastructures que les entreprises hébergent et maintiennent dans leurs propres installations par opposition à l'exécution à distance sur des serveurs dans le cloud.

En installant et en exécutant des logiciels sur du matériel situé dans les locaux de l'entreprise, les équipes IT ont un accès physique aux données et peuvent contrôler directement la configuration et la sécurité de l'infrastructure.

Intelligence Artificielle :

L'IA est un domaine de recherche visant à reproduire par des systèmes artificiels les diverses capacités cognitives de l'être humain (et plus largement les comportements "intelligents" du vivant), en particulier en ce qui concerne leur capacités à résoudre des problèmes complexes.

IoT :

L'internet des objets (Internet of Things : IoT) désigne l'ensemble des objets du quotidien dotés d'une capacité intelligente. Ils fournissent un flux de données en continu, et sont connectés à un système de gestion centralisé afin de partager et d'échanger des données, mais aussi d'interagir avec d'autres objets connectés. Ils peuvent ainsi prendre des décisions de manière autonomes et effectuer des tâches utiles en se basant sur des conditions prédéfinies.

M

M2M :

Le M2M ou machine-to-machine fait référence aux nombreuses interactions entre les objets connectés : smartphones, TV, objets du quotidien (montre, frigo...).

Machine learning :

Le machine learning ou "apprentissage automatisé" est un sous-domaine de l'IA qui regroupe les approches statistiques. Il existe plusieurs méthodologies d'apprentissage automatisé identifiées par la typologie de problèmes qu'elles sont à même de résoudre : l'apprentissage supervisé, non supervisé, semi-supervisé et par renforcement

Métadonnées :

Ensemble d'informations qualifiant et donnant des précisions sur une valeur dans une base de données. Si l'on prend l'exemple de la reconnaissance d'image, le processus de labellisation permet d'obtenir une description pour chacune des images de la banque de données. Il existe de nombreux outils permettant de mettre en place ces dictionnaires de la donnée (Collibra, Informatica...).

Modèle algorithmique :

Un modèle algorithmique est un ensemble d'éléments (flux de données entrants, séries d'algorithmes et flux de données sortants) permettant de répondre à la question/au problème posé.

N

Neurone artificiel :

—

Le neurone artificiel (autrement appelé neurone formel) est un modèle mathématique et informatique inspiré du neurone biologique.

Neurones profonds :

—

Les neurones artificiels profonds sont également appelés neurones fonctionnels. On parle de réseaux de neurones profonds lorsqu'ils sont regroupés en plusieurs "couches", définissant ainsi la profondeur du réseau. (cf. [définition deep learning](#)).

P

PAAS (Platform as a Service, en anglais) :

—

Le PaaS est un modèle de service de cloud computing. L'infrastructure physique et le système d'exploitation sont fournis par la solution cloud. Le client peut installer ses propres applications si besoin. Google App Engine, Microsoft Azure ou encore AWS Elastic Beanstalk en sont des exemples. (cf. [schéma Les 4 unités d'échelle du cloud computing p.63](#)).

Paramètres d'un modèle :

—

Les paramètres d'un modèle contiennent entre autre les hyper-paramètres : ce sont des coefficients et indications de structure du modèle.

PétaFLOP :

—

Un pétaFLOP est l'unité de mesure de la puissance de calcul des microprocesseurs.

R

Ramp-up test :

—

Le Ramp-up test est utilisé lorsqu'on souhaite tester la performance d'une application. L'un des aspects les plus critiques des tests de performances est la capacité à imiter le trafic entrant et sortant d'une application. En appliquant le Ramp-up, on prépare l'application à répondre à tous types de situation.

Rétro-engineering :

—

Dans l'univers de la programmation, le rétro-engineering est utilisé lorsque l'on souhaite comprendre le fonctionnement interne d'un programme. C'est une activité qui consiste à décomposer les différentes étapes d'un programme pour analyser leur fonction et leurs interactions. Elle sert notamment à lutter contre l'effet "black-box" d'un modèle algorithmique.

RGPD :

—

Le règlement général sur la protection des données est entré en vigueur en Europe depuis mai 2018. Il encadre le traitement des données à caractère personnel par les organisations afin d'assurer le respect de la vie privée et le contrôle des informations personnelles partagées.

S

SAAS :

—

Le SAAS (software as a service) est une typologie d'infrastructure. C'est une couche d'abstraction du cloud computing permettant de saisir les données sur un logiciel, les données sont ensuite transformées à distance via une interface logicielle sans qu'un ordinateur soit impliqué.

Système multi-agents :

—

Un système multi-agents (SMA) désigne les interactions entre plusieurs agents - entités computationnelles autonomes (exemples: programme informatique ou robot) - dans un environnement commun. Ces derniers permettent de réduire la complexité de la résolution d'un problème en divisant le savoir-faire en sous-ensembles et en coordonnant l'activité de chacun des agents. Dans le cadre d'applications d'IA, on parle d'intelligence artificielle distribuée.

T

Tensorflow :

—

Framework de machine learning open source développé par Google.

TPUs :

—

Les TPUs (Tensor Processing Units) sont une catégorie de processeurs principalement utilisés pour la phase de réponse des réseaux de neurones.

V

Variables cibles :

—

Les variable(s) cible(s) se sont la ou les informations attendues en "sortie" du modèle : c'est ce que l'on cherche à obtenir.

Variables explicatives :

—

Les variables explicatives sont les données d'entrée sur lesquelles le modèle se base pour déduire la variable cible. Aussi appelées features, elles viennent peupler la base d'apprentissage du modèle. Elles sont créées par combinaison des données initiales pour les "rendre comestibles" pour l'algorithme.

Webliographie,

p.52

Défense et aérospatial :

—

<https://www.nasa.gov/feature/goddard/2017/nasa-explores-artificial-intelligence-for-space-communications>

<https://www.airbus.com/newsroom/news/en/2016/12/Artificial-Intelligence.html>

<https://www.airbus.com/newsroom/news/en/2016/09/one-atlas-airbus-defence-and-space-launches-the-worlds.html>

<http://geospatial-solutions.com/airbus-selects-google-cloud-for-one-atlas-basemap/>

<http://decryptageo.fr/one-atlas-airbus-defence-and-space-lance-la-bibliotheque-dimages-satellite-les-plus-recentes-au-monde/>

https://www.accenture.com/t20180517T074106Z_w_us-en/_acnmedia/PDF-78/Accenture-Seizing-Digital-Opportunity-Aerospace-Defense.pdf

p.53

Agronomie :

—

<https://www.satellitetoday.com/imagery-and-sensing/2018/10/11/how-will-precision-agriculture-feed-the-world/>

<https://globenewswire.com/news-release/2018/09/21/1574215/0/en/Global-Precision-Agriculture-Market-2018-2025-Focus-on-Solution-Hardware-System-Software-and-Service-and-Application-Irrigation-Field-Monitoring-Spraying-Fertilization-Planting-and.html>

<https://www.blog.google/technology/ai/using-tensorflow-keep-farmers-happy-and-cows-healthy/>

<https://bensohillbio.com/benson-hill-news-insights/page/5/>

<https://cen.acs.org/business/agriculture/Benson-Hill-Becks-unveil-corn/96/i21>

<http://www.libelium.com/smart-wine-libeliums-iot-technology-allows-predictive-control-of-vineyards-in-the-pago-ayles-winery-spain/>

<https://www.forbes.com/sites/bernardmarr/2018/03/09/the-incredible-ways-john-deere-is-using-artificial-intelligence-to-transform-farming/#5e5bfb330dfc>

<https://www.usinenouvelle.com/article/meme-les-vaches-peuvent-etre-connectees.N676219>

p.54

Produits de consommation et distribution :

—

<http://www.e-marketing.fr/Thematique/ia-1250/Breves/Comment-intelligence-artificielle-investit-elle-magasins-312743.htm>

<https://www.frenchweb.fr/cas-reel-searskmart-engager-100-des-clients-mobile-en-magasin-sans-appli-avec-intelligence-artificielle/321589>

<https://www.alizila.com/alibaba-launches-big-data-anti-counterfeiting-alliance/>

<https://www.entrupy.com/>

p.55

Energie et télécommunications :

—

<https://energyfactor.exxonmobil.com/news/mit-collaboration/>

<https://www.groupe-estia.fr/chez-direct-energie-lanalyse-predictive-eclaire-comportement-clients/>

<https://www.esilv.fr/smart-grid-vers-troisieme-revolution-industrielle-solution-a-crise-energetique/>

<https://www.futura-sciences.com/planete/actualites/developpement-durable-metroscope-logiciel-intelligent-auscultation-industrielle-70776/>

<https://www.ge.com/digital/manufacturing-execution-systems>

p.56

Hôtellerie et tourisme :

—

<https://www.lechotouristique.com/article/quand-un-chatbot-imité-un-conseiller-voyage>

<https://www.tourhebdo.com/actualites/technologie/le-tourisme-craque-pour-les-chatbots-447063.php>

<https://businesstravel.accorhotels.com/travelstyle-blog/digital-traveler/gerer-un-voyage-a-ere-des-chatbots.fr.shtml>

<https://medium.com/hijiffy/top-3-chatbots-that-are-changing-the-travel-industry-d325082c50b8>

https://www.washingtonpost.com/lifestyle/travel/talking-up-travel-industry-chatbots/2018/05/16/dc705816-5564-11e8-abd8-265bd07a9859_story.html?noredirect=on&utm_term=.38810924c790

<https://www.mytravelresearch.com/chatbots-and-the-travel-industry/>

<https://mezi.com/>

<https://techcrunch.com/2016/11/15/ai-powered-virtual-assistant-mezi-pivots-to-focus-on-travel/>

<https://www.tom.travel/2017/10/27/bot-daccorhotels-desormais-disponible-google-assistant/>

<https://www.avvio.com/booking-engine/>

<https://www.altexsoft.com/blog/datascience/machine-learning-redefines-revenue-management-and-dynamic-pricing-in-hotel-industry/>

<https://futurism.com/automated-video-surveillance-streamline-airport-experience/>

p.57

Mobilité et transports :

—

<https://www.tesla.com/>

<https://roboticsandautomationnews.com/2018/05/17/gm-uses-ai-based-generative-design-technology-for-reducing-vehicle-weight/17289/>

<https://www.cbtnews.com/what-is-ai-and-how-can-it-benefit-my-dealership/>

<https://hbr.org/2017/05/how-harley-davidson-used-predictive-analytics-to-increase-new-york-sales-leads-by-2930>

<https://www.konux.com/>

<https://eng.uber.com/machine-learning/>

p.58

Santé :

—

<https://www.crainsdetroit.com/sponsored-content/big-data-disruption-how-big-data-analytics-can-impact-health-care>

<https://www.telegraph.co.uk/technology/2018/10/15/googles-ai-can-spot-breast-cancer-better-humans/>

<http://scopeblog.stanford.edu/2013/09/27/big-data-big-finds-clinical-trial-for-deadly-lung-cancer-launched-by-stanford-study/>

<https://techcrunch.com/2013/04/05/asthmapolis-wants-to-hack-the-inhaler-and-help-26-million-americans-better-track-and-manage-their-asthma/>

p.59

Médias et divertissement :

—

<https://medium.com/netflix-techblog/meson-workflow-orchestration-for-netflix-recommendations-fc932625c1d9>

<https://medium.com/syncedreview/ai-in-the-media-and-entertainment-industry-1ad4b2b701b8>

p.60

Education et secteur public :

—

https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

<https://news.microsoft.com/en-sg/2016/07/12/singapore-to-explore-next-generation-digital-government-services-with-conversations-as-a-platform-proof-of-concept/>

<https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay-refugees-claim-asylum-legal-aid>

<https://www.technologyreview.com/s/610181/what-it-felt-like-to-visit-the-most-tech-centric-olympics-ever/>

<https://govinsider.asia/innovation/japan-trials-ai-for-parliament-use/>

p.61

Banques et services financiers :

—

<https://dssg.uchicago.edu/project/optimizing-the-quality-and-delivery-of-city-emergency-medical-services/>

<https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf>

<https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay-refugees-claim-asylum-legal-aid>

<https://www.latribune.fr/entreprises-finance/banques-finance/ia-watson-est-un-investissement-rentable-au-credit-mutuel-769495.html>

<https://newsroom.mastercard.com/eu/press-releases/mastercard-rolls-out-artificial-intelligence-across-its-global-network/>

<https://www.shift-technology.com/>

<http://www.larevuedudigital.com/societe-generale-multiplie-par-4-sa-detection-du-churn-grace-a-li-a/>

Infos,

Tous droits réservés. Cet ouvrage ne peut en aucune manière être reproduit en tout ou partie, sous quelque forme que ce soit ou encore par des moyens mécaniques ou électroniques, y compris le stockage de données et leur retransmission par voie informatique sans autorisation d'Artefact.

La citation des marques est faite sans aucun but publicitaire. Les erreurs ou les omissions involontaires qui auraient pu subsister dans cet ouvrage malgré les soins et les contrôles d'Artefact ne sauraient engager leur responsabilité.

Impression

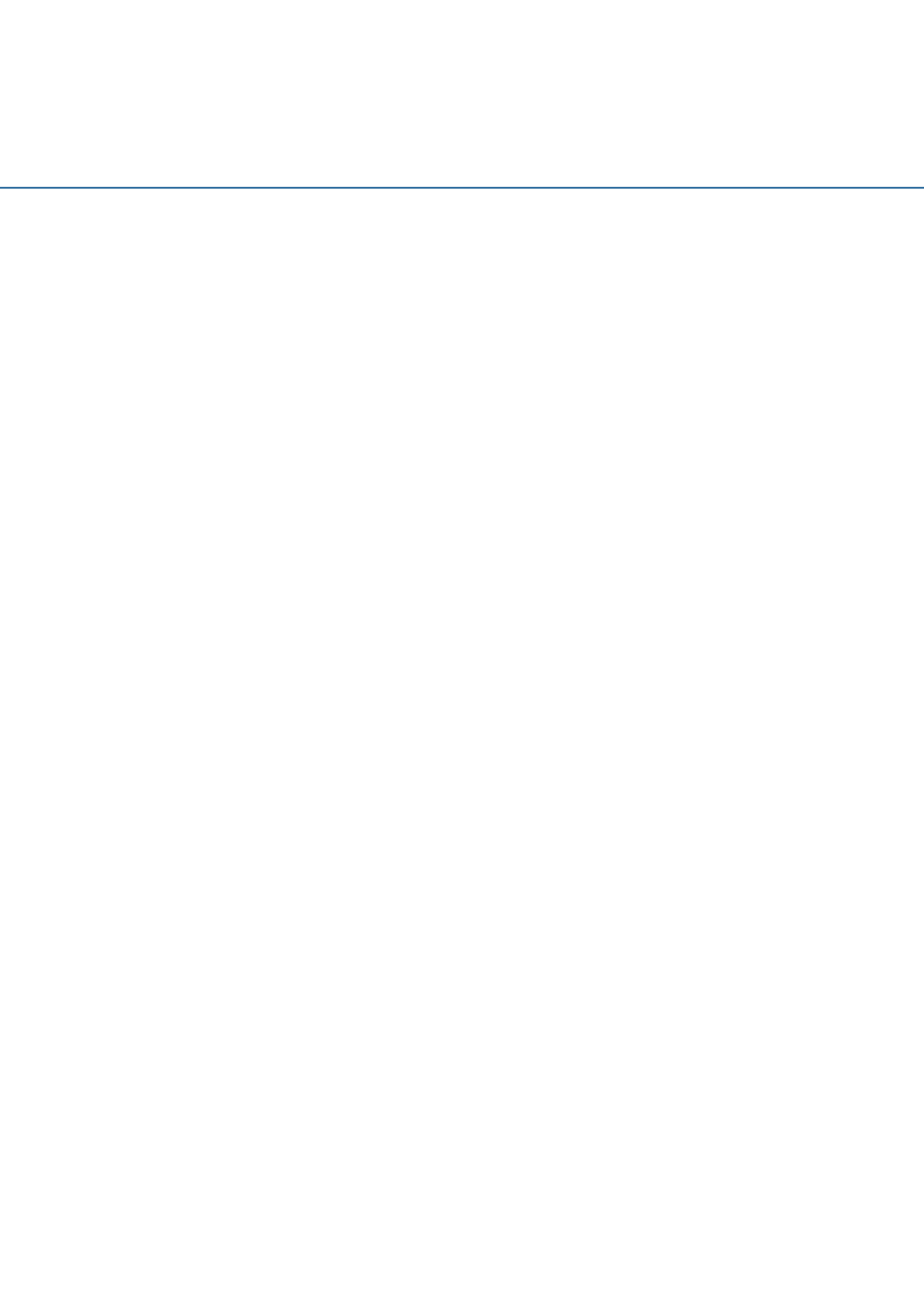
Corlet Imprimeur

Rédaction

Justine Nerce
Solène Marie
Pascal Coggia

Conception graphique/maquette

Malvina Alves
Annabel Deschamps





ARTEFACT