



Part of Dun & Bradstreet's
Worldwide Network



LIVRE BLANC

**Maîtriser la donnée pour maîtriser l'IA :
le choix stratégique des IA privées**

SOMMAIRE

INTRODUCTION	3
L'INTELLIGENCE ARTIFICIELLE : QU'EST-CE QUE C'EST ?	4
LES OBJECTIFS ET PERSPECTIVES DE L'IA	5
QU'EST CE QU'UN MODELE DE LANGAGE ?	5
LES MODÈLES DE LANGAGE SELON LEUR GOUVERNANCE ET LE NIVEAU DE CONTRÔLE DE LA DONNÉE	7
QUI UTILISE LES DONNÉES DANS L'ENTREPRISE ?	9
CRÉER UNE IA PRIVÉE EN ENTREPRISE : DU CHOIX STRATÉGIQUE À L'EXÉCUTION	10
LA DONNÉE AU CŒUR DU PROCESSUS DE CRÉATION D'UNE IA PRIVÉE	11
UNE DONNÉE DE QUALITÉ : PILIER DE LA CONFORMITÉ RÉGLEMENTAIRE D'UNE IA	14
COMMENT DISPOSER DE LA MEILLEURE DONNÉE POUR SON IA ?	15
CONCLUSION	16

Introduction

L'intelligence artificielle est sur toutes les lèvres. Elle fascine autant qu'elle effraie, et promet de révolutionner notre manière de travailler, de décider et même de penser. Elle s'impose aujourd'hui comme un levier majeur de transformation de la société et en particulier des entreprises.

Qu'il s'agisse d'automatiser des tâches, d'anticiper des risques, d'améliorer la relation client ou de soutenir la décision, l'IA s'intègre progressivement à tous les niveaux de l'organisation. D'un côté, on trouve les modèles d'IA dits ouverts - proposés par les grands acteurs du cloud et accessibles via Internet, que chacun peut utiliser. De l'autre, des approches plus cadrées se multiplient, avec des entreprises qui développent ou adaptent leurs propres modèles, fondés sur leurs données internes et gouvernés selon leurs règles: on les appelle les IA privées.

Quel est le meilleur choix pour une entreprise ?
Est-ce bien prudent de confier ses données à une IA ?

Chez Altares, nous sommes convaincus que la maîtrise de la donnée conditionne la maîtrise de l'IA. Dans ce livre blanc, nous allons passer en revue l'histoire de l'IA, les différents types d'IA et modèles de langage qui s'offrent à nous, et réfléchir aux options pour s'assurer la meilleure qualité de donnée possible pour alimenter une IA.



L'Intelligence Artificielle : qu'est-ce que c'est ?

Lorsque l'on parle d'IA aujourd'hui, tout le monde pense instantanément à ChatGPT, l'agent conversationnel qui peut écrire et corriger à notre place, mais l'IA ne se limite pas à ce domaine.

L'intelligence artificielle désigne une technologie visant à concevoir des systèmes capables de réaliser des tâches habituellement réservées aux capacités cognitives humaines : percevoir, comprendre, raisonner, apprendre, interagir ou agir de manière autonome dans un environnement donné. Et cette ambition ne date pas d'hier.

L'HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE (IA)

L'idée d'intelligence artificielle n'est pas aussi récente qu'on pourrait le penser.

Au fil des siècles, les philosophes et les savants ont cherché à comprendre si la pensée humaine pouvait être traduite en mécanismes logiques, et ont tenté de concevoir des calculs, formules et machines à cet effet. Mais c'est au XIXe siècle que les concepts d'IA commencent à se préciser, et que la fiction rejoint la science :

la formulation mathématique du processus de raisonnement humain et de logique jette les bases mathématiques de la pensée calculable.

Au XXe siècle, l'ordinateur programmable concrétise ces intuitions : en 1956, John McCarthy définit l'intelligence artificielle comme discipline scientifique. Après des phases d'euphorie et de stagnation, l'IA s'impose dans les années 2000 avec le machine learning et l'explosion des données. Depuis 2010, la puissance de calcul et le deep learning propulsent une nouvelle génération d'IA capables de percevoir, anticiper et interagir : reconnaissance d'images, traitement du langage, assistants intelligents.

Aujourd'hui, nous sommes entrés dans une nouvelle ère : celle où la machine ne se contente plus d'exécuter, mais devient un collaborateur.



Les objectifs et perspectives de l'IA

L'IA repose sur l'hypothèse que la pensée humaine peut être mécanisée. Partant de ce postulat, les ingénieurs en IA développent des algorithmes et modèles informatiques visant à reproduire les fonctions cognitives via, principalement, la communication en langage naturel, la vision artificielle, et la simulation du raisonnement et de la compréhension.

Automatiser des tâches, améliorer la prise de décisions, renforcer la sécurité et la conformité : dans le monde professionnel, l'IA est de plus en plus intégrée, et représente une formidable opportunité et un outil de performance et de transformation, à condition d'être précautionneux sur les données qui sont utilisées.

Et pour faire son choix parmi ce qui existe sur le marché, il faut avant tout bien comprendre quelles options s'offrent à nous. Penchons-nous sur les différents types d'IA, et dans quels cas ils peuvent être mis à profit.

Qu'est-ce qu'un modèle de langage ?

Un modèle de langage (LLM, Large Language Model) est un type d'intelligence artificielle entraîné sur d'immenses volumes de données textuelles, qui lui servent à comprendre et à générer du texte. En pratique, il apprend à prédire le mot suivant dans une phrase, ce qui lui permet de :

- répondre à des questions (notamment dans le cadre de l'automatisation du support client) ;
- rédiger des textes (de plus en plus sophistiqués et à la fois naturels !) ;
- traduire (avec subtilité) ;
- résumer en quelques secondes ;
- aider à la prise de décision avec des analyses et recommandations ;
- coder, etc.



LLM, SLM ET MACHINE LEARNING : QUEL CHOIX POUR QUELLES DONNÉES ET QUELS USAGES ?

Tous les modèles d'IA ne servent pas les mêmes objectifs. Dans une IA privée, le choix dépend surtout des données disponibles, de leur gouvernance et des usages métiers.

LLM (LARGE LANGUAGE MODELS) : très puissants et polyvalents, entraînés sur d'immenses volumes de données, ils comprennent le langage de manière générale. Leur déploiement est coûteux et complexe à contrôler pour des cas métiers précis.

Dans une IA privée, le plus important n'est pas la taille du modèle, mais la qualité et la maîtrise des données. Un modèle ciblé et alimenté par des données fiables produira des résultats plus pertinents qu'un modèle volumineux nourri de données peu contrôlées.

SLM (SMALL LANGUAGE MODELS) : plus compacts et spécialisés, ils offrent d'excellentes performances sur des périmètres ciblés, à condition de s'appuyer sur des données fiables, structurées et contextualisées. Leur taille facilite le contrôle, la traçabilité et la conformité.

Mais jusqu'où peut-on vraiment confier ses données à un modèle que l'on ne contrôle pas ?

MACHINE LEARNING CLASSIQUE : efficace pour le scoring, la prédiction ou la classification à partir de données structurées, il reste pertinent pour des usages métiers précis.



Les modèles de langage selon leur gouvernance et le niveau de contrôle de la donnée

Au-delà des usages, un autre critère fondamental distingue les modèles d'IA : leur mode de gouvernance. Celle-ci permet d'évaluer le niveau de contrôle, de sécurité et de souveraineté associé à chaque modèle.

L'IA OUVERTE

Les IA ouvertes sont accessibles via des plateformes publiques. Elles offrent une grande facilité d'usage et une mise en œuvre rapide, mais impliquent un faible niveau de contrôle sur les données transmises et sur leur usage. Elles sont adaptées à des cas non sensibles, mais présentent des limites en matière de confidentialité, de conformité et de souveraineté.

L'IA HYBRIDE (RAG – RETRIEVAL AUGMENTED GENERATION)

L'IA hybride, grâce au RAG, va chercher l'information à la demande dans des bases internes et des bases externes de confiance, sans intégrer ces données dans son apprentissage.

Cette approche permet de garantir la fraîcheur des informations, réduire les hallucinations, et conserver un contrôle strict sur les sources utilisées. Pour le chef de projet IA, le RAG présente plusieurs avantages décisifs. Il garantit la fraîcheur des informations, puisque les réponses reposent sur des données mises à jour en temps réel. Le RAG permet aussi un haut niveau de maîtrise et de traçabilité, car chaque source interrogée peut être identifiée, documentée et auditée. Enfin, les risques d'hallucinations sont considérablement réduits grâce à des sources factuelles et vérifiées.

Dans une logique d'IA privée, le RAG devient un outil stratégique de gouvernance de la donnée. Il permet d'ouvrir l'IA à des référentiels externes de qualité tout en conservant un contrôle strict sur ce qui est consulté.

L'IA PRIVÉE

L'IA privée repose sur des modèles entraînés ou adaptés exclusivement à partir de données internes, gouvernées et sécurisées. L'entreprise conserve la maîtrise complète du modèle, de ses sources et de ses usages.

Elle est particulièrement adaptée aux contextes exigeants qui utilisent des données sensibles, avec des enjeux réglementaires forts, et des besoins en fiabilité et en traçabilité élevés.

Les principaux types d'intelligence artificielle en entreprise

On distingue plusieurs formes d'IA, mais concentrons-nous sur les trois qui concernent surtout les usages en entreprise :

L'IA PRÉDICTIVE / ANALYTIQUE

L'IA prédictive analyse des données historiques et actuelles pour anticiper des comportements ou des événements futurs. Elle transforme des volumes importants de données en informations exploitables pour la prise de décision.

Elle est utilisée pour :

- le scoring de risque (clients, fournisseurs, partenaires) ;
- la détection de fraude ou d'anomalies ;
- l'anticipation des défaillances, de perte de clients ou des tensions sur la chaîne d'approvisionnement ;
- l'analyse des tendances de marché.

C'est aujourd'hui la forme d'IA la plus mature et la plus intégrée dans les systèmes décisionnels des entreprises.

L'IA GÉNÉRATIVE (GENAI)

L'IA générative ne se contente pas d'analyser, elle produit de nouveaux contenus. En s'appuyant sur des modèles de langage entraînés sur d'immenses volumes de données, elle est capable de générer du texte, du code, des images ou des synthèses complexes.

Elle alimente notamment :

- les assistants conversationnels ;
- la rédaction automatisée ;
- la synthèse documentaire ;
- l'aide à la prise de décision.

Mais elle est à manier avec prudence, car si ses capacités sont spectaculaires et très plébiscitées, elle pose aussi de nouveaux enjeux de gouvernance, notamment sur la qualité des données utilisées et le risque d'hallucinations qui ont déjà coûté gros à certaines structures imprudentes.

Un autre élément essentiel doit être pris en compte par les entreprises : contrairement à de nombreux systèmes de machine learning traditionnels, l'IA générative est par nature non déterministe.

Là où un modèle de machine learning classique nécessite généralement un volume très important de données pour apprendre à identifier un élément précis (par exemple un numéro de facture), un modèle génératif peut produire un résultat à partir d'un nombre limité de documents.

Toutefois, la réponse produite n'est jamais garantie et des erreurs peuvent se produire, ce qui impose la mise en place de mécanismes de contrôle renforcés, dont une supervision humaine.

L'IA doit ainsi être envisagée comme un outil d'assistance à la décision, et non comme un système autonome infaillible. La combinaison d'une IA performante, de données fiables et d'un contrôle humain demeure aujourd'hui la meilleure garantie de fiabilité.

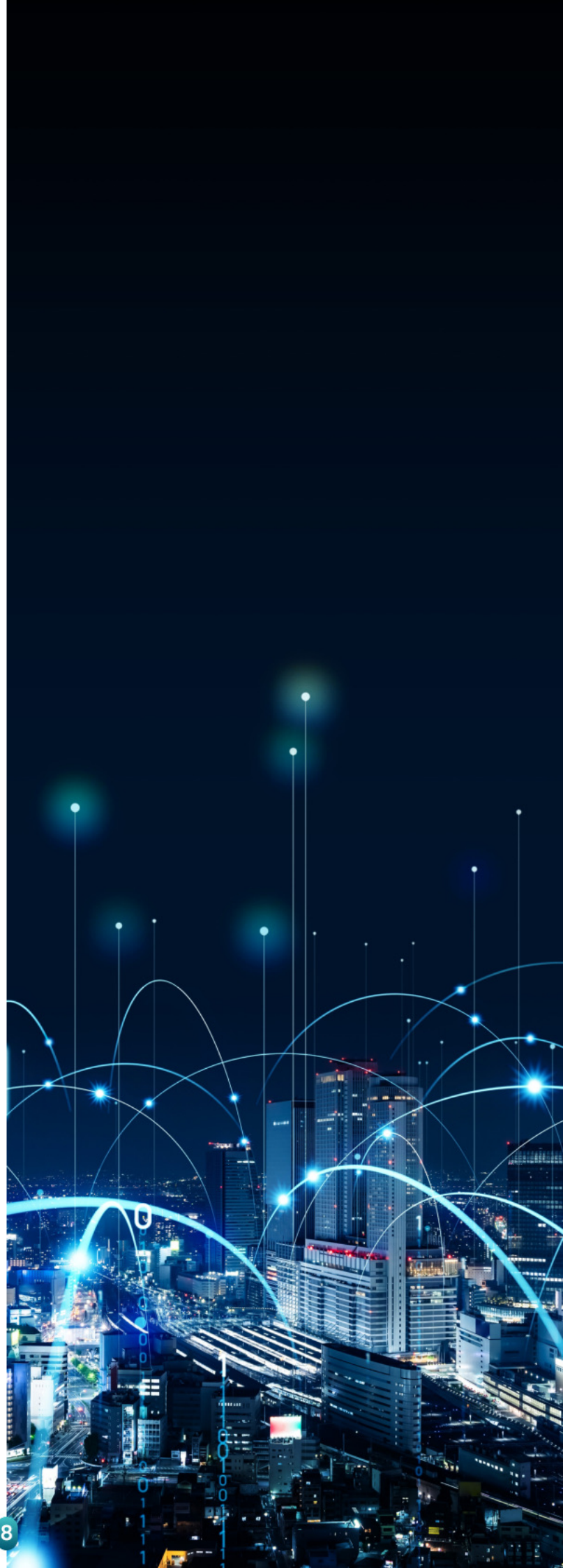
L'IA AGENTIQUE OU OPÉRATIONNELLE

Ces modèles ne se limitent plus à répondre ou à produire du contenu : ils agissent.

Connectée aux outils métiers (CRM, ERP, Excel, plateformes marketing, etc.), l'IA agentique peut :

- exécuter des chaînes d'actions complexes ;
- automatiser des processus de bout en bout ;
- orchestrer des décisions opérationnelles en temps réel.

Cette capacité d'action de l'IA renforce encore l'exigence de fiabilité, de traçabilité et de maîtrise de la donnée, car une erreur peut engendrer des impacts opérationnels non négligeables.



Qui utilise la donnée dans l'entreprise ?

L'ensemble du personnel d'une entreprise utilise certes la donnée, mais pas de la même manière, ni pour les mêmes finalités. Chaque département mobilise des sources et des indicateurs spécifiques selon ses objectifs.

Direction et management

Leur objectif est de piloter la stratégie et la performance globale de l'entreprise. Ils mobilisent des données financières, commerciales et opérationnelles issues des ERP, des outils de pilotage et des tableaux de bord BI (Business Intelligence). Ces informations leur permettent de suivre les indicateurs clés (chiffre d'affaires, marge, croissance, performance par activité), d'arbitrer les investissements et d'orienter les décisions stratégiques.

IT, data & sécurité

Leur rôle est de rendre la donnée accessible, fiable et sécurisée. Ils s'appuient sur des données systèmes, des logs, des métadonnées, des bases utilisateurs et des référentiels MDM (Master Data Management) pour garantir la cohérence et la sécurité du patrimoine informationnel.

Juridique, conformité & risques

Leur objectif est de maîtriser les obligations réglementaires et les risques associés. Ils utilisent des bases d'entreprises (Altaires, D&B, Infogreffe), des données contractuelles, des données réglementaires et des modèles de scoring de risque pour assurer la conformité et anticiper les vulnérabilités.

Ressources humaines

Leur mission est de gérer les talents et d'améliorer la performance sociale. Ils exploitent des données issues du SIRH (Système d'Information des Ressources Humaines), de la paie, des plateformes e-learning, des feedbacks employés et des entretiens annuels.

Opérations & production

Ces équipes visent à garantir la qualité, la continuité et la performance opérationnelle. Elles utilisent des données de capteurs, les systèmes ERP (Enterprise Resource Planning, système de gestion des processus), PLM (Product Lifecycle Management, système de gestion du cycle de vie d'un produit) et MES (Manufacturing Execution System, système de gestion de la production en atelier), ainsi que des bases techniques assurant le suivi des processus industriels.

Sales, marketing & relation client

Leur objectif est de générer du revenu et fidéliser les clients. Ils exploitent des données issues du CRM, des comportements digitaux (web, e-mail), des données d'entreprises (firmographie, liens capitalistiques), des données d'usage produit ou encore des indicateurs de satisfaction.

Nous le voyons, la donnée est au cœur des processus de tous les départements de l'entreprise, des organes de direction jusqu'aux partenaires externes, en passant par la chaîne de production et les équipes support. Il est essentiel que les données circulent avec fluidité et sans perte ou déformation, pour garantir que l'ensemble de l'organisation parle le même langage, dispose des mêmes éléments et les retrouve dans l'IA privée de leur entreprise.

Utilisateurs de la donnée dans une entreprise

Direction



R&D



Partenaires / écosystème externe



DSI - IT



Service client / support



Supply chain



Ressources Humaines



Finance



Achats



Juridique



Data team



Sales



Marketing

Créer une IA privée en entreprise : du choix stratégique à l'exécution

LA DÉCISION STRATÉGIQUE

De plus en plus d'entreprises envisagent une IA privée afin de limiter les risques liés à la circulation de données sensibles. Ce choix n'est jamais neutre : il s'inscrit à la croisée des exigences réglementaires (RGPD, AI Act), des enjeux de confidentialité, du risque d'erreur et de la nécessité d'automatiser des processus critiques.

L'entreprise doit alors arbitrer entre des modèles ouverts, performants mais difficiles à maîtriser, et une IA privée, plus contrôlée mais plus exigeante en termes de moyens, de gouvernance et d'exécution. Ce choix stratégique appelle rapidement une question clé : qui pilote concrètement le projet ?

LE PILOTAGE DU PROJET IA

La mise en œuvre d'une IA privée repose sur le chef de projet IA. Il ne s'agit pas d'un expert technique isolé, mais d'un rôle de coordination chargé de faire le lien entre stratégie, données, métiers et contraintes réglementaires.

Ce rôle requiert une compréhension solide des fondamentaux de l'IA (machine learning, modèles de langage, architectures de données), sans pour autant être purement scientifique. Son rôle est de transformer des besoins métiers en choix concrets, compatibles avec les contraintes de données et de conformité.

Il combine ainsi compétences en gestion de projet, connaissance des enjeux métier et maîtrise de la réglementation. Son rôle est de coordonner les parties prenantes (data, IT, sécurité, juridique, métiers) et d'anticiper les risques dès les premières phases du projet. Cette fonction implique notamment de structurer les étapes du projet et d'aligner les choix technologiques avec les priorités de l'entreprise.

Avec ce pilotage en place, le projet peut alors entrer dans sa phase opérationnelle.

POURQUOI DES ENTREPRISES CRÉENT-ELLES LEUR PROPRE IA PRIVÉE ?

Avant de se lancer, la question à se poser est la suivante : pourquoi choisir une IA privée plutôt qu'un modèle ouvert ? La première raison est le contrôle. Une IA privée permet de maîtriser les données utilisées, la façon dont les modèles sont exploités et les réponses produites. Elle s'impose dès lors que l'entreprise manipule des données sensibles, soumises à des exigences fortes de confidentialité, de conformité ou de souveraineté.

Ce choix permet aussi d'adapter l'IA aux réalités métiers, de l'intégrer aux outils internes (CRM, intranet, support) et de limiter la dépendance à des solutions externes dont les coûts et les règles d'usage échappent souvent à l'entreprise. Pour certaines organisations, c'est également un moyen de valoriser un patrimoine de données unique, difficilement exploitable dans des modèles ouverts.

Mais une IA privée a un revers : elle suppose un effort constant sur la donnée. Maintenance, mises à jour, qualité et cohérence des référentiels deviennent des prérequis. La donnée n'est plus un simple carburant, elle devient le cœur du projet. C'est à ce stade que le pilotage prend tout son sens : avant de parler modèle ou algorithme, l'entreprise doit s'assurer de disposer de bases de données fiables, complètes et à jour.



La donnée au cœur du processus de création d'une IA privée

Le chef de projet connaît une règle fondamentale : une IA fiable ne peut fonctionner qu'avec des données fiables. Des informations erronées, incomplètes ou obsolètes peuvent vite créer des erreurs, semer le doute et nuire à la confiance des utilisateurs. À l'inverse, des données solides et bien vérifiées deviennent le moteur qui fait vraiment fonctionner l'IA. Des exemples récents le montrent clairement : le cabinet Deloitte Australie a remis un rapport au ministère du travail truffé de citations inventées et d'erreurs, entraînant un remboursement coûteux et des conséquences sur sa réputation. En France, près de 500 cas juridiques ont révélé des résultats "hallucinés" fournis par des IA, faussant dossiers et analyses.

Pour éviter ce genre de pièges, la première mission du chef de projet est simple mais cruciale : vérifier que les données respectent cinq critères fondamentaux.

LES CRITÈRES D'UNE DONNÉE FIABLE POUR L'IA

Pour qu'une donnée alimente efficacement vos projets d'intelligence artificielle, elle doit respecter cinq exigences essentielles. Elle doit avant tout être exacte, c'est-à-dire refléter fidèlement la réalité, sans approximation. Elle doit aussi être complète / exhaustive, autrement dit contenir l'ensemble des informations nécessaires à une exploitation pertinente. Sa fraîcheur doit être garantie, elle doit être tenue à jour en continu pour rester pertinente. Sa traçabilité est aussi un critère primordial, vous devez pouvoir retracer son origine et ses évolutions. Enfin, la qualité d'une donnée passe aussi par son accessibilité : elle doit être disponible pour les bonnes personnes et les bons modèles, tout en respectant les règles de sécurité et de conformité.

En utilisant vos propres bases de données dans votre propre IA, vous augmentez vos chances de générer des résultats fiables et exempts d'erreurs.

CRITÈRES D'UNE DONNÉE FIABLE

Exactitude

Complétude

Fraîcheur

Accessibilité

Traçabilité

COMMENT METTRE À PROFIT VOTRE DONNÉE DANS VOTRE IA ?

La donnée, même fiable, n'a de valeur que si elle est exploitée intelligemment. Dans un projet d'IA privée, le traitement et la gouvernance des données ne se limitent pas à leur stockage : il s'agit aussi de les sécuriser, structurer et transformer en un véritable actif stratégique. Penchons-nous sur les étapes de cette gestion des données et leur usage dans un modèle d'IA.

Construire et sécuriser son patrimoine de données

Avant de penser modèle ou algorithme, il faut d'abord bâtir un patrimoine de données solide et protégé. C'est une responsabilité centrale que le chef de projet supervise.

Cela implique de collecter des données propriétaires (CRM, documents internes, FAQ, code source, etc.) pour nourrir l'IA avec des données réellement différenciantes. Vient ensuite la phase critique de nettoyage, structuration et anonymisation des données afin d'éliminer doublons et biais, tout en préservant la confidentialité. L'entreprise doit aussi garantir un hébergement souverain à sa donnée, c'est-à-dire exclure tout partage en dehors de l'entreprise, opter pour un stockage sur serveurs internes, un cloud privé ou une infrastructure certifiée et conforme.

Une donnée bien gouvernée constitue le premier rempart contre les risques de fuite, de biais ou de non-conformité réglementaire. Une fois cette base maîtrisée, elle peut être exploitée intelligemment dans l'IA.

DE LA DATA À L'ALGORITHME : ENTRAÎNER ET CONTRÔLER LE MODÈLE

Une fois la base de données prête et fiable, le projet passe à l'étape clé : l'entraînement du modèle. Le chef de projet accompagne chaque étape pour s'assurer que tout se déroule correctement.

Concrètement, on part souvent d'un modèle existant (généralement open source) et on le réentraîne avec les données propres à l'entreprise. Cela permet de profiter de la robustesse d'un modèle déjà éprouvé tout en l'adaptant aux spécificités du métier.

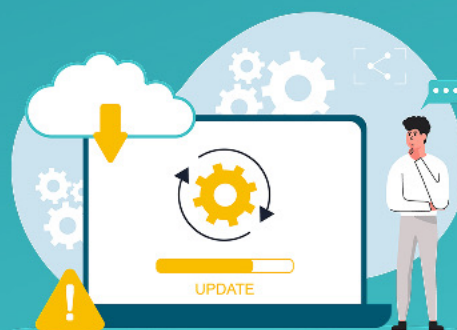
Il est aussi essentiel de protéger les paramètres du modèle issus de l'entraînement, pour éviter toute utilisation externe ou modification non autorisée. Une fois stabilisé, le modèle doit être utilisé dans un cadre sécurisé : via une API interne ou un portail réservé, pour garantir un usage contrôlé et conforme aux règles de l'entreprise.

Cette approche permet de créer un outil unique, qui comprend le langage de l'entreprise et applique ses règles. C'est tout l'intérêt d'une IA privée : maîtriser ses données, préserver leur confidentialité et obtenir des analyses fiables, adaptées aux besoins réels de l'organisation.

LES APPLICATIONS CONCRÈTES DE L'IA EN ENTREPRISE

Avec une IA entraînée et sécurisée, le chef de projet peut enfin déployer et monétiser cette solution. Les analyses prédictives de l'IA offrent une multitude d'applications concrètes pour aider l'entreprise à progresser et à prendre de meilleures décisions, pour peu qu'elles se basent sur des données de haute qualité.

- **Prévention des risques** : identifier les clients, fournisseurs ou partenaires présentant des signes de fragilité avant qu'ils ne deviennent critiques.
- **Optimisation financière** : ajuster les politiques de crédit, de paiement ou d'assurance selon les niveaux de risque observés.
- **Fidélisation client** : détecter les signaux de désengagement et recommander des actions ciblées pour renforcer la rétention.
- **Anticipation des évolutions de marché** : repérer les tendances, évaluer les opportunités de diversification ou de croissance.
- **Détection de fraude ou de non-conformité** : surveiller les transactions et les comportements anormaux pour renforcer la sécurité.
- **Pilotage de la supply chain** : suivre la santé financière des fournisseurs stratégiques pour garantir la continuité d'activité.
- **Automatisation intelligente** : déléguer aux algorithmes les décisions répétitives et à faible enjeu, tout en concentrant les équipes sur les choix stratégiques.



LES RISQUES D'UNE DONNÉE DÉFAILLANTE DANS L'IA

A contrario, une IA nourrie avec des données imparfaites n'est pas seulement moins performante : elle peut mettre en péril l'activité de l'entreprise.

Les principaux risques d'une IA alimentée par des mauvaises données :



DÉCISIONS BIAISÉES :

De mauvaises données en entrée produisent de mauvais résultats en sortie (garbage in, garbage out). Dans le domaine financier, un scoring inexact peut écarter de bons clients ou accorder des conditions à des profils risqués. Dans la santé, une mauvaise donnée d'entrée peut fausser un diagnostic ou une alerte prédictive. Ces biais ne se voient pas toujours immédiatement, mais leurs effets s'accumulent avec le temps.



PERTE DE CONFIANCE :

Des résultats incohérents freinent l'adoption et l'usage des solutions d'IA. Si l'utilisateur de l'IA constate des manquements, ou pire, des erreurs dans les résultats, il risque de prendre à juste titre ses distances avec les outils qui lui paraîtront peu fiables. Et s'il ne vérifie pas les résultats, il se retrouvera confronté ultérieurement aux conséquences des erreurs non détectées. Dans les deux cas, la confiance est entamée.



SURCOÛTS :

Corriger, réentraîner et réexécuter les modèles consomment du temps, du budget et des ressources. Une large part du cycle de développement d'un modèle d'IA est consacrée à la préparation des données. Plus ces données sont de mauvaise qualité, plus leur exploitation demande du temps et des ressources.



NON-CONFORMITÉ :

RGPD, IA Act, LCB-FT et autres réglementations imposent une donnée fiable, traçable et maîtrisée. La conformité passe par une gouvernance rigoureuse de la donnée, une mise à jour continue des référentiels et une parfaite transparence sur l'origine et l'usage des informations.

Une IA mal alimentée devient une machine à erreurs ; à l'inverse, une IA privée fondée sur des données maîtrisées devient un facteur de confiance et de performance durable.

Une donnée de qualité : pilier de la conformité réglementaire d'une IA

Pour le chef de projet, la conformité n'est pas seulement une étape à valider à la fin du projet. C'est un fil conducteur tout au long du développement de l'IA. Il doit rester attentif à la provenance des données et à leur utilisation, en respectant le RGPD, l'IA Act et la directive NIS2.

- **Anticipation des risques** : Le chef de projet doit prévenir les sanctions financières lourdes et les ruptures opérationnelles qui surviendraient si un audit révélait l'usage de données non conformes.
- **Instaurer la confiance** : En s'appuyant sur des données fiables et vérifiées (comme le D-U-N-S® Number), il montre aux régulateurs et aux partenaires que l'entreprise prend ses obligations au sérieux.
- **Garantie de continuité** : Une donnée traçable permet de justifier chaque décision de l'IA à tout moment, et de s'assurer que les opérations de l'entreprise peuvent se dérouler sans encombre.

Focus sur les architectes de la donnée dans l'entreprise

Les data scientists (ainsi que les data analysts et data engineers) sont des acteurs centraux dans la chaîne de valeur de la donnée, mais ils ne sont pas des « utilisateurs finaux » au sens métier : ils en sont les architectes, traducteurs et valorisateurs.

Data Engineer

Le data engineer conçoit, automatise et met en production les systèmes de collecte et de traitement des données pour les rendre exploitables par les équipes métiers et d'analyse.

- Collecte, nettoyage, normalisation, intégration des données (internes et externes)
- Construction des pipelines (ETL, API, streaming)
- Mise en place des data lakes, data warehouses, MDM
- Gestion de la qualité et de la gouvernance des données

Data Analyst

Le data analyst exploite les données de l'entreprise pour créer des analyses, tableaux de bord et modèles statistiques aidant les services opérationnels à prendre des décisions éclairées. Il rend les données lisibles et exploitables.

- Création de dashboards, indicateurs, visualisations
- Analyses descriptives et exploratoires
- Aide à la décision pour les équipes métier (marketing, sales, finance, RH, etc.)

Data Scientist

Le data scientist conçoit et met en œuvre des modèles de machine learning pour répondre aux besoins métier, en structurant et analysant les données nécessaires à leur développement et leur déploiement.

- Modélisation, machine learning, IA, scoring, recommandations
- Analyse prédictive : churn, risque client, fraude, maintenance prédictive, etc.
- Traitement du langage naturel (NLP), computer vision, etc.

CDO (Chief Data Officer)

Le chief data officer pilote la gouvernance, la qualité et la conformité des données pour soutenir la transformation digitale de l'entreprise. Il encadre et orchestre l'ensemble du cycle de vie de la donnée.

- Définit la stratégie data, les rôles, les droits d'accès
- Garantit la conformité (RGPD, sécurité, souveraineté)
- Pilote la « culture data » dans toute l'entreprise

Ensemble, ces experts garantissent que la donnée circule, reste fiable et exploitable à chaque étape de son cycle de vie. Leur coordination conditionne la capacité de l'entreprise à disposer d'un socle de données solide, indispensable à toute stratégie d'intelligence artificielle.

Comment disposer de la meilleure donnée pour son IA ?

Le chef de projet sait qu'une IA performante ne repose pas sur une accumulation massive et désordonnée de fichiers, mais sur un ensemble de données organisé et fiable. Son rôle est de sélectionner les informations qui apporteront une réelle valeur ajoutée sans compromettre la sécurité.

Plutôt que de chercher à collecter toujours plus de volume, il se concentre sur la fiabilisation et l'enrichissement des données déjà présentes dans le patrimoine interne de l'entreprise. Pour cela, il agit en trois étapes : consolider le socle interne, ouvrir de manière maîtrisée vers l'extérieur et tirer parti des retours d'expérience du modèle.

STRUCTURER ET FIABILISER LES DONNÉES INTERNES

Les données internes, provenant du CRM, de l'ERP ou des intranets, sont le cœur du projet. Le chef de projet veille à transformer ces informations brutes en un ensemble cohérent et facile à exploiter.

- **Nettoyage et cohérence** : Il organise le tri et la correction des doublons, anomalies et formats incohérents, car des données incorrectes peuvent vite fausser les résultats.
- **Sécurité et accès** : Il fixe les règles pour protéger les informations et contrôler qui peut y accéder, en s'assurant que seules les données pertinentes sont utilisées.
- **Clarté métier** : En annotant et en classant les informations, il s'assure que tout le monde, et le système, comprend bien le contexte propre à l'entreprise.

ENRICHIR LES DONNÉES AVEC DES SOURCES PÉRIPHÉRIQUES ET EXTERNES

Pour que l'IA ne tourne pas en vase clos, le chef de projet doit l'ouvrir intelligemment à des sources complémentaires.

- **Données de terrain** : Il intègre les informations issues du support client (FAQ, tickets, retours terrain) pour mieux comprendre le quotidien et les interactions de l'entreprise.

- **Sources fiables** : Il se sert de référentiels externes fiables (bases SIREN, partenaires certifiés) tout en appliquant les mêmes règles de rigueur et de traçabilité que pour les données internes.
- **Mise à jour continue** : Il met en place des connecteurs et des API (comme Dataxess ou Direct+) pour que les informations soient actualisées en continu, évitant ainsi que le système travaille sur des données obsolètes.

EXPLOITER LES DONNÉES GÉNÉRÉES PAR L'IA ELLE-MÊME

Dans une stratégie d'IA privée, le travail du chef de projet ne s'arrête pas au déploiement : il utilise l'IA pour **générer de la nouvelle connaissance**. Chaque interaction, chaque log et chaque feedback utilisateur produit une donnée précieuse.

- **Mesure de performance** : Ces données permettent au chef de projet de mesurer l'efficacité réelle de l'outil et d'identifier les axes d'amélioration ou les biais éventuels.
- **Boucle d'apprentissage** : Il met en place des processus pour tirer parti des réussites et des erreurs de l'outil afin d'ajuster progressivement son fonctionnement.
- **Auto-enrichissement** : Sous sa direction, l'IA devient un levier qui améliore la qualité globale du patrimoine de données de l'entreprise, créant ainsi un système qui gagne en pertinence et en valeur au fil du temps.

MAINTENIR LA DONNÉE À JOUR DANS LES SYSTÈMES

Développer sa propre IA demande aux entreprises un investissement permanent et sur le long terme, afin de maintenir sa pertinence et son efficacité. Cela nécessite une mise à jour constante des données de l'entreprise à partir de sources sûres. Pour ce faire, des solutions API et des connecteurs tels que Dataxess, Business Add-On et Direct+ permettent d'enrichir et maintenir vos données dans vos systèmes, grâce à la base mondiale d'Altaires-D&B.

Conclusion

Construire une IA efficace en entreprise ne consiste pas à accumuler massivement des données, mais à organiser et structurer celles dont on dispose. Dans ce cadre, chaque information doit être documentée, qualifiée, tracée et mise à jour en continu pour devenir un véritable actif stratégique.

C'est cette rigueur, bien plus que la simple puissance de calcul ou la quantité de fichiers stockés, qui distingue aujourd'hui les organisations capables de créer une intelligence artificielle de confiance. Le chef de projet IA, en tant que coordinateur, porte la responsabilité de transformer une base de données brute en un levier concret pour la prise de décision.

Le défi pour les entreprises n'est plus seulement de "faire de l'intelligence artificielle", mais de garantir que cette technologie ne devienne pas une source de risques réputationnels ou opérationnels. Comme nous l'avons vu, une donnée défaillante mène inévitablement à des erreurs coûteuses ou à des "hallucinations" qui brisent la confiance des utilisateurs. À l'inverse, une IA privée, alimentée par un socle de données solides et fiables, offre une protection inégalée pour le patrimoine informationnel de l'entreprise.

C'est ici qu'interviennent les fondations mêmes de votre projet. Altares accompagne les entreprises dans cette exigence de qualité en mettant à leur disposition la plus grande base de données BtoB au monde. Continuellement enrichie et vérifiée par des experts, cette source de référence permet d'alimenter votre IA avec des données fiables.

Cette fiabilité fait d'Altares un acteur clé pour bâtir des IA qui ne sont pas seulement performantes, mais aussi souveraines et parfaitement conformes aux exigences réglementaires actuelles, comme le RGPD ou l'AI Act. En choisissant l'excellence de la donnée, le Chef de projet IA ne se contente pas de déployer un outil : il prépare l'entreprise à naviguer avec assurance dans l'ère de l'intelligence artificielle responsable.

