

# PLATEFORME SÉMANTIQUE POUR LE DÉPLOIEMENT DE L'IA ET DES DONNÉES À GRANDE ÉCHELLE DANS L'ENTREPRISE

INTRODUCTION AU FRAMEWORK TRAIDA (TRANSFORMATIVE AI AND DATA SOLUTIONS)

04 Mars 2024

Ce rapport en français est la traduction du rapport en version originale anglaise,  
également disponible sur notre site : [www.engage-meta.com](http://www.engage-meta.com).

**Mots-clés** : système entreprise, approche systémique, plateforme sémantique, base de données orientée graphes de connaissances, gouvernance des données, gouvernance de l'IA, IA responsable.

Dans ce rapport, nous désignons par "système entreprise" l'ensemble du système permettant à l'entreprise de réaliser ses activités. Il comprend le système d'information (applications, bases de données, processus automatisés, infrastructure), l'organisation (acteurs, rôles, connaissances et processus non informatisés, valeurs, politique de gestion des ressources humaines) et les algorithmes d'IA.

## INTRODUCTION

Ce rapport présente une approche globale pour la mise en œuvre à grande échelle de l'intelligence artificielle (IA) au sein des entreprises, en mettant l'accent à la fois sur les perspectives stratégiques et techniques. Notre objectif est d'établir une architecture et un cadre adaptés au déploiement expansif de l'IA dans l'ensemble du système entreprise d'une organisation, plutôt que de se pencher sur des applications verticales de l'IA. Ce livre blanc vise à servir de guide pour les cadres et les décideurs, sans imposer de connaissances techniques obligatoires, pour initier des tâches opérationnelles telles que le cadrage stratégique autour de l'IA responsable, comprendre son impact sur les attitudes des équipes, les besoins en formation, l'outillage logiciel, et l'identification des premiers cas d'application. Bien que nous nous concentrons sur l'architecture du système entreprise, notre but est aussi d'augmenter votre compréhension du domaine technique afin de saisir les tendances clés qui façonnent l'évolution de votre entreprise à l'ère de l'IA. Vous devriez ainsi collecter quelques connaissances utiles pour aligner votre système entreprise avec le pouvoir transformatif de l'IA et de ses données.

## AUTEURS

Le principal auteur est **Pierre Bonnet**, expert en architecture d'entreprise (*Enterprise Architecture*), gouvernance des données (*Master Data Management*) et accompagnement des entreprises dans l'intégration de nouvelles technologies. Il est également le fondateur de la communauté Engage-Meta, un groupe de réflexion axé sur le développement de pratiques pour renforcer l'accumulation des connaissances, ainsi que les pratiques autour des données et de l'IA. Le co-auteur est le **Professeur Nguyễn Xuân Hoài**, expert en intelligence artificielle et Directeur de l'AI Academy Vietnam.

- Pierre Bonnet – HLFL Consulting – [pierre.bonnet@hlfl-consulting.com](mailto:pierre.bonnet@hlfl-consulting.com)
- Nguyễn Xuân Hoài – AI Academy - [nxhoai@aiaacademy.edu.vn](mailto:nxhoai@aiaacademy.edu.vn) , (<https://aiaacademy.edu.vn>)



Le contenu de cette publication est libre d'utilisation, publié sous une licence Creative Commons, ce qui signifie qu'il est en open-source. Vous êtes libre de réutiliser les concepts, les textes et les illustrations comme vous le souhaitez, à condition de citer la source, qui est la communauté "Engage-meta".

# Table des matières

<b>INTRODUCTION .....</b>	<b>1</b>
Auteurs .....	1
<b>APPORTS ET RISQUES DE L'IA .....</b>	<b>3</b>
Utilisation extensive.....	3
Potentiel significatif de problèmes.....	3
<b>BARRIÈRES .....</b>	<b>3</b>
Qualité insuffisante des données en provenance des silos.....	3
Manque d'accumulation personnelle et collective des connaissances .....	4
Manque d'alignement entre le métier et la technique.....	4
<b>PLATEFORME SÉMANTIQUE .....</b>	<b>4</b>
IA responsable .....	4
Deuxième cerveau.....	8
Architecture complète.....	8
<b>MISE EN OEUVRE (TRAIDA FRAMEWORK).....</b>	<b>9</b>
Principes .....	9
Cartes techniques .....	9
Cartes de la gouvernance.....	10
Cartes du métier.....	10
Procédure .....	11
1. Evaluation .....	11
2. Alignement avec le métier.....	12
3. Portefeuille des projets de transformation .....	13
<b>QUELLES SONT LES PREMIÈRES ÉTAPES? .....</b>	<b>14</b>
<b>CONCLUSION.....</b>	<b>15</b>
<b>ANNEXE : BASE DE DONNÉES ORIENTÉES GRAPHS .....</b>	<b>16</b>

# APPORTS ET RISQUES DE L'IA

## UTILISATION EXTENSIVE

L'IA est un ensemble d'outils informatiques qui s'améliorent vite et deviennent de plus en plus conviviaux. **Sa barrière à l'entrée a été brisée avec la sortie de ChatGPT en novembre 2022.** Aujourd'hui, il n'est plus nécessaire d'être informaticien pour utiliser l'IA. Ses contributions sont exponentielles et s'intègrent dans notre quotidien ainsi que dans celui des entreprises. Dans ce contexte, il est naturel pour les décideurs de s'inquiéter de l'avenir de leur entreprise s'ils venaient à manquer la prochaine vague de l'IA. Être dépassé par des concurrents mieux équipés en intelligence artificielle représente un risque existentiel immédiat pour toutes les entreprises.

## POTENTIEL SIGNIFICATIF DE PROBLEMES

Cependant, comme l'a dit Milton Friedman, « il n'y a pas de repas gratuit ». Le côté positif des contributions exponentielles de l'IA a également un côté obscur. Ce sont les risques exponentiels générés par son mauvais usage. Tant que l'entreprise déploie ces technologies sur quelques applications verticales, les dysfonctionnements et le manque de fiabilité de l'IA sont compensés par l'organisation. **L'intelligence humaine sait encore faire face aux échecs de l'IA, mais seulement jusqu'à un certain seuil.** Il arrive un moment où ces applications verticales de l'IA déconnectées les unes des autres, génèrent une masse de problèmes que l'organisation ne peut plus résoudre. Dans cette situation, l'entreprise changera de cap pour réduire son utilisation de l'IA et prendra le risque d'être dépassée par des acteurs du marché plus efficaces. Elle subira alors une croissance exponentielle de problèmes nuisant à la motivation des employés. Il est probable qu'en l'absence d'anticipation d'un déploiement de l'IA à grande échelle, un chaos technique et humain s'ensuive. **Si vos succès initiaux en IA masquent votre préparation à son déploiement à grande échelle, vous êtes dans la situation de quelqu'un qui saute d'un avion sans parachute tout en se disant « jusqu'ici tout va bien ».** Bien sûr, cette préparation n'est pas facile et nécessite beaucoup d'expertise. C'est pourquoi nous avons rédigé ce document introductif. Nous proposons des mesures pour un déploiement à grande échelle de l'IA. Il s'agit d'une architecture pour une « IA responsable » et un cadre de mise en œuvre appelé TRAIDA (Transformative AI and Data solutions). Ils fournissent les meilleures pratiques pour réussir un déploiement progressif et opérationnel.

## BARRIÈRES

### QUALITE INSUFFISANTE DES DONNEES EN PROVENANCE DES SILOS

Les grandes entreprises possèdent des systèmes informatiques complexes avec de nombreuses applications et bases de données hétérogènes. Ces systèmes sont verticalisés autour de silos fonctionnels et organisationnels. Les défauts de qualité sont alors multiples : duplication des données générant des erreurs ; processus métier entrant en collision avec les limites des silos ; et solutions d'intégration technique complexes, coûteuses et fragiles (ETL, EAI, MDM, virtualisation des données). Peu d'entreprises entreprennent la tâche de réviser leurs silos pour développer des applications propres et des bases de données organisées par domaine d'activité. Bien que l'industrie informatique propose des technologies pour y parvenir, telles que l'architecture orientée services (*Service Oriented Architecture* – SOA) et l'architecture de données décentralisée (*data mesh*), leur mise en œuvre reste délicate. Ce genre de refonte impose une transformation sur plusieurs années et mobilise des budgets importants avec des compétences difficiles à trouver. Face au poids de cette transformation et pour éviter de remettre en question tous les silos, les entreprises privilégient des solutions préservant les existants. C'est là que les outils de gestion des données maîtres et de référence, d'intégration des données clients, de virtualisation des données et d'*operational data store* (ODS) entrent en jeu. Ils impliquent la mise en œuvre de couches d'agrégation de données et de référentiels de données transversaux aux silos. Ils unifient alors, plus ou moins bien, les données pour augmenter la pertinence de leur gouvernance et leur qualité. L'entreprise conserve ainsi ses silos et améliore partiellement la gestion de ses données avec des solutions imparfaites qui préservent le système existant. **Pourtant, avec l'IA, cette gouvernance approximative des données n'est plus acceptable.** En l'absence de données de qualité à travers tout le système entreprise, les applications IA ne fonctionnent pas correctement, et il n'est pas toujours possible de compenser les erreurs qu'elles génèrent. De plus, **l'application de l'IA dans des opérations métier en silo ou verticales ne produit pas un retour sur investissement élevé.**

Afin de maximiser l'avantage de l'IA pour la transformation de l'activité, chaque entreprise devrait appliquer l'IA de manière large, transversale et coordonnée à travers de multiples opérations métier. Dans le reste de ce document, vous verrez comment la "**plateforme sémantique**" offre une solution pour améliorer profondément la qualité des données à travers tout le système entreprise, sans imposer une refonte des silos. Elle devient alors le socle pour le déploiement d'une IA fiable et responsable.

## MANQUE D'ACCUMULATION PERSONNELLE ET COLLECTIVE DES CONNAISSANCES

Pour les experts en informatique, l'accélération et l'étendue de la technologie de l'information ne permettent plus à un individu seul de tout connaître (a). Le domaine d'expertise de l'ingénieur en informatique se réduit ainsi à quelques outils verticaux demandés par le marché à une certaine époque. Il est normal que les écoles forment à ces besoins, cependant **il est nécessaire de mettre davantage l'accent sur les compétences transversales telles que la modélisation des données et des processus, l'architecture d'entreprise ou la gouvernance des systèmes complexes**. Bien qu'elles soient de moins en moins enseignées, ces disciplines sont essentielles pour la réussite des projets de transformation avec l'IA et les données. Ces projets sont par nature transversaux au sein des entreprises, avec des implications significatives pour l'architecture et la gouvernance du système. **En termes d'activité métier, les connaissances sont mal formalisées (b)**. En effet, les solutions informatiques s'intéressent aux données structurées nécessaires au fonctionnement des logiciels et pas assez aux connaissances des individus. La majorité des connaissances reste dans le cerveau des acteurs avec peu de partage collectif. De plus, le manque d'appétit pour l'écriture entrave la formalisation des savoirs alors qu'elle doit être renforcée pour les besoins d'entraînement de l'IA.

## MANQUE D'ALIGNEMENT ENTRE LE METIER ET LA TECHNIQUE

Cette fragmentation des connaissances en informatique (a) et le manque de formalisation des connaissances métier (b) ne facilitent pas l'alignement entre le business et la technologie. **Il y a un besoin et une attente pour plus d'architectes expérimentés possédant une culture technique large sur le système d'information, ainsi que plus d'experts des métiers capables de documenter formellement les connaissances**. Il manque également des pratiques pour documenter et suivre l'alignement des métiers avec la technologie, particulièrement pour les projets de transformation.

# PLATEFORME SÉMANTIQUE

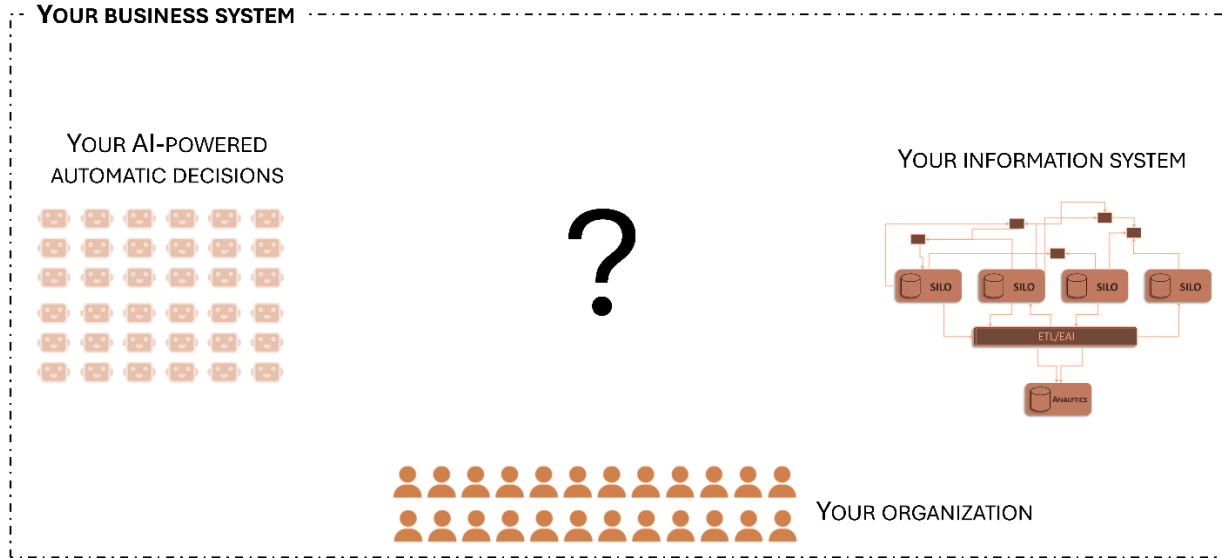
## IA RESPONSABLE

Les obstacles que nous venons de décrire peuvent être levés pour cibler une IA fiable et responsable à condition que :

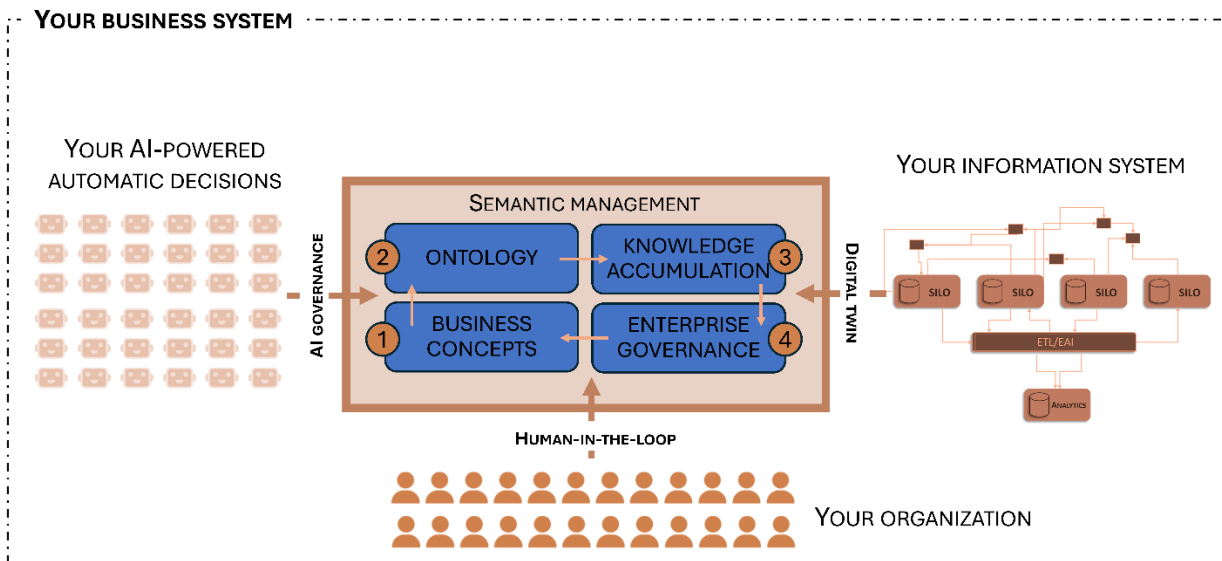
1. Les défauts de qualité des données soient résolus profondément sans imposer une refonte systématique des silos ;
2. Les connaissances des acteurs soient capitalisées et utilisées pour entraîner les applications d'IA ;
3. L'alignement entre le business et la technique soit mieux formalisé.

Pour y parvenir, il est nécessaire de s'appuyer sur une architecture afin d'intégrer le système d'information, l'organisation et l'IA à l'échelle de l'entreprise ; il faut aussi un cadre pour accumuler les pratiques nécessaires aux déploiements des projets de transformation avec les données et l'IA. Cette architecture repose sur une « plateforme sémantique » ; et le cadre pour l'accumulation des pratiques est basé sur le framework TRAIDA (Transformative AI and Data solutions).

Par conséquent, pour déployer avec succès l'IA à grande échelle, ces trois domaines du système entreprise doivent être coordonnés (a, b, c) :

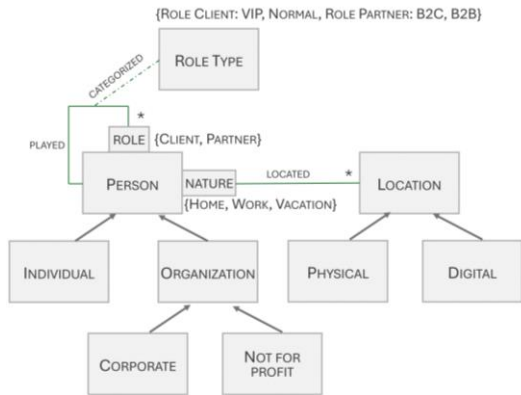


Premièrement, sur le côté droit de la figure se trouve le **système d'information (a)**, illustré ici avec ses silos. Il représente les applications, les bases de données, les fichiers et les processus automatisés. Sur le côté gauche de la figure se trouve **l'intelligence artificielle (b)**, représentée comme un portefeuille d'algorithmes et d'outils intelligents capables de prendre des décisions automatisées. La contribution de l'IA réside beaucoup dans sa capacité à améliorer l'automatisation de la prise de décision. Pour fonctionner, l'IA a besoin de sources de données fiables provenant du système d'information et doit interagir avec le troisième domaine présenté dans cette figure, à savoir **l'organisation (c)**. Le point d'interrogation au centre de l'illustration ne peut rester sans réponse apportée par une infrastructure qui couvre tous les cas d'utilisation de l'IA. Sans elle, il y aurait des connexions dans toutes les directions entre les trois domaines (a, b,c), avec des risques pour la qualité et la traçabilité des résultats. Le plat de spaghetti qui existe souvent avec la gestion des données au sein des silos risquerait alors d'être encore amplifiée au niveau de l'IA. Par conséquent, une infrastructure centrale doit être proposée pour coordonner l'ensemble du système. Examinons ses principes avec la figure suivante.



Notre vision repose sur une « **plateforme sémantique** » qui assure la fonction de coordination des trois domaines formés par le système d'information, l'organisation et l'intelligence artificielle. Cette plateforme fournit un glossaire de concepts métier, qui sont les éléments de données clés et les objets métier utilisés pour l'exécution des opérations dans l'ensemble de l'entreprise. Ils doivent être définis de manière unifiée (glossaire d'entreprise) et sont utilisés pour construire l'ontologie à partir de laquelle les algorithmes d'IA puisent les données.

L'ontologie est une instance du modèle sémantique d'information qui inclut des règles fondamentales d'intégrité, telles que le cycle de vie des concepts métier. **Cette ontologie est également le catalyseur pour accumuler les connaissances individuelles et collectives détenues par les acteurs au sein de l'organisation.** Ces connaissances sont cruciales pour entraîner les algorithmes d'IA. L'objectif est de faciliter la collecte et la classification des connaissances individuelles et collectives par l'écriture. Enfin, la plateforme bénéficie d'une gouvernance au niveau de l'entreprise pour garantir sa pertinence dans le temps.



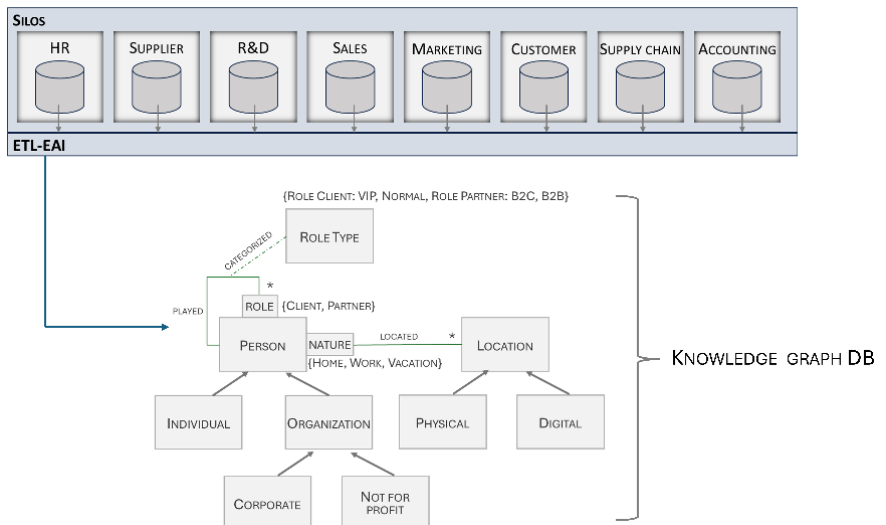
La plateforme sémantique agit comme un jumeau numérique du système d'information, masquant les détails des applications et des bases de données à la fois pour les algorithmes d'IA et pour l'organisation. Elle permet de corriger les défauts de qualité dans les systèmes existants. Elle est essentielle pour déployer l'IA à grande échelle et garantir sa bonne gouvernance (IA fiable et responsable).

Un exemple de modèle sémantique est présenté dans la figure opposée. Une instance du modèle sémantique (ontologie) décrit le concept métier d'une « Personne ». Vous pouvez observer différentes taxonomies, c'est-à-dire des classifications comme les types de personne tels que Individuel, Organisationnel, etc., ou les types de rôles comprenant Client, Partenaire et Employé.

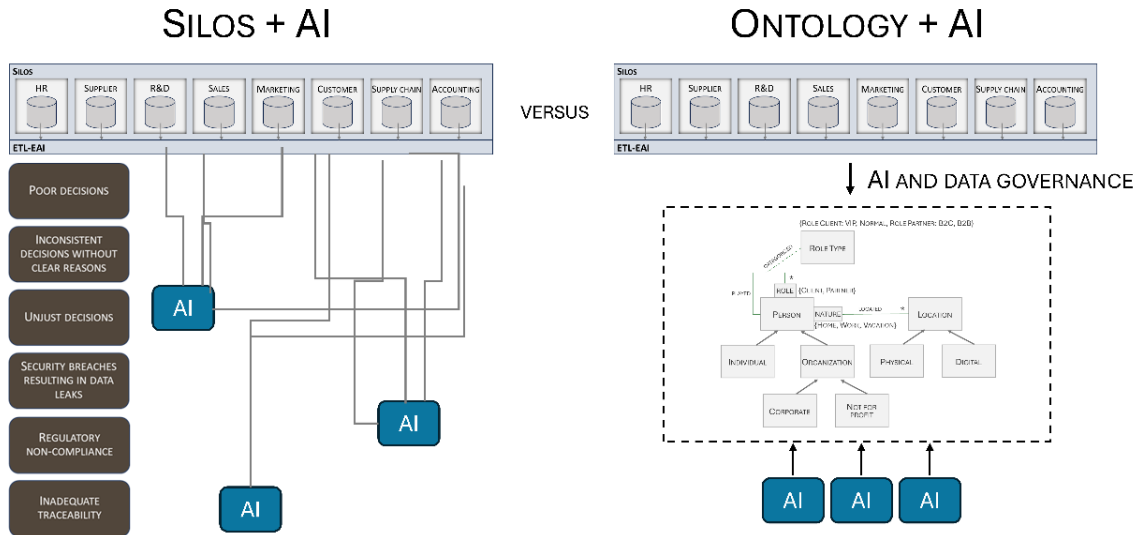
L'ontologie est mise en œuvre dans un référentiel de données transversal aux bases de données des silos. **Une bonne pratique consiste à utiliser la technologie des bases de données orientées graphes de connaissances (voir annexe).**

La synchronisation des données entre les silos et l'ontologie nécessite une bonne maîtrise de la gestion des flux de données et des technologies telles que l'ETL et l'EAI. Cette synchronisation des données est un sujet courant en informatique.

Lorsque l'ontologie est en lecture seule, la synchronisation est relativement simple à mettre en œuvre, mais elle devient plus complexe lorsque l'ontologie est également mise à jour.



Dans la partie gauche de la figure ci-dessous, les algorithmes d'IA sont directement connectés aux sources de données à l'intérieur des silos. Cela conduit à la propagation des défauts de qualité des données dans les algorithmes d'IA. De plus, l'absence de centralisation empêche l'établissement d'une gouvernance globale pour les applications d'IA et les opérations. Les problèmes sont alors multiples : la qualité des décisions avec l'IA est limitée en raison de l'adhésion aux silos ; des décisions incohérentes entre les silos émergent, conduisant à un manque de confiance en l'IA. En outre, l'accès non autorisé à certaines données au sein des silos conduit à des violations de la sécurité, à une mauvaise application des réglementations et, en fin de compte, à un manque de traçabilité dans les opérations d'IA.



Dans la partie droite, les algorithmes d'IA s'appuient sur une ontologie qui masque les défauts de qualité des données dans les silos et dissimule la complexité d'accès aux bases de données au sein de ces mêmes silos. L'ontologie permet une gouvernance unifiée de l'IA et des données associées. Les défauts de qualité mentionnés sont alors éliminés.

Si vous démarrez vos projets d'IA sans la plateforme sémantique, vous devez faire attention à ne pas aller trop loin ou vous risquez de créer une complexité et des fragilités difficiles à corriger.

**Une IA avec des silos et des défauts de qualité des données ne peut pas être considérée comme une 'IA responsable'. La plateforme sémantique facilite la construction et le déploiement d'une IA responsable à travers tout le système entreprise. Par conséquent, il est vivement recommandé de mettre en œuvre une première version de l'ontologie dès les premiers algorithmes d'IA.**

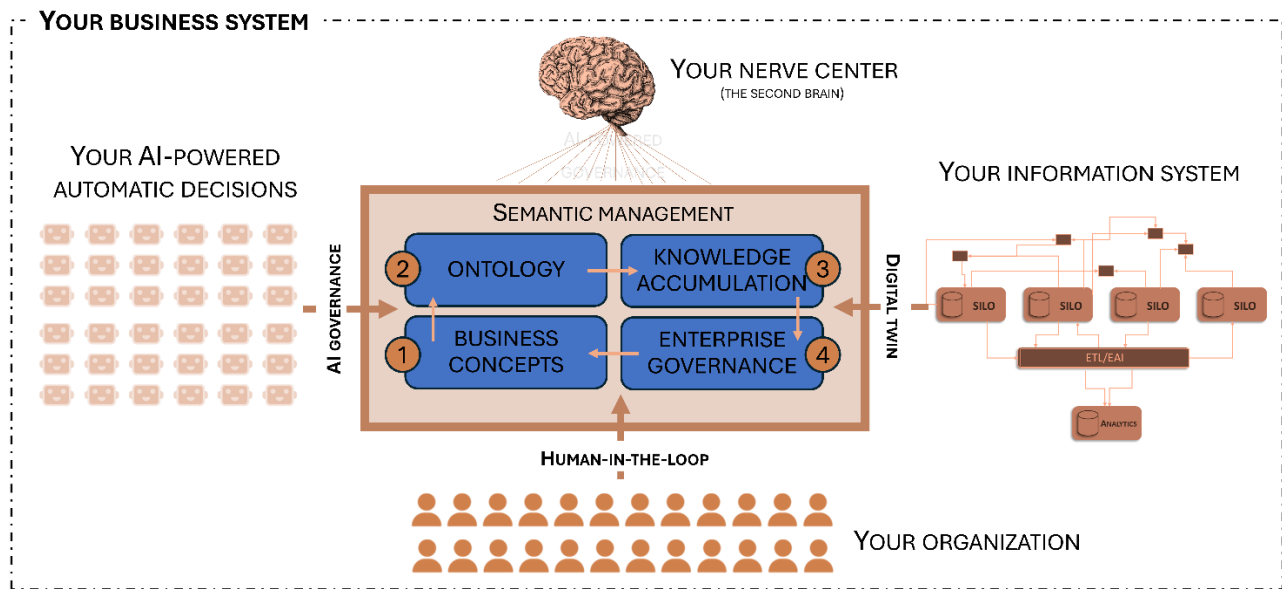
## DEUXIEME CERVEAU

Notre vision intègre un concept final dont le rôle est de superviser le comportement de la plateforme sémantique. En assurant la qualité de son exécution, c'est également la gouvernance de tout le système entreprise qui est mis sous contrôle.

En effet, la plateforme sémantique peut également être une source d'erreurs, tout comme les algorithmes de prise de décision de l'IA, sans oublier les défauts de qualité et les dysfonctionnements provenant du système d'information. Sans l'utilisation de l'IA, il est impossible de coder des algorithmes assez puissants et flexibles pour superviser le comportement de toutes les applications, données et processus. Aujourd'hui, avec l'IA, cela devient possible. Le principe est de fournir à une IA autant de documentation que possible sur le système d'information, les données, les algorithmes, l'organisation et les réglementations. Cela peut inclure des spécifications, des guides d'utilisation, des schémas de bases de données, des descriptions et captures d'écran d'interface utilisateur, des descriptions vocales, des comptes-rendus de réunions, etc. Cette IA observe le comportement du système pour détecter des actions atypiques, erronées ou frauduleuses. Elle peut également découvrir des améliorations qu'un humain ne serait pas capable de détecter. On l'appelle la '**gouvernance alimentée par l'IA**', le centre nerveux du système entreprise, également appelé le 'deuxième cerveau'. **Plus le portefeuille d'algorithmes d'IA pour la prise de décision automatique est riche, plus l'intelligence globale du système entreprise augmente, nécessitant alors un système de gouvernance alimenté par l'IA.**

## ARCHITECTURE COMPLETE

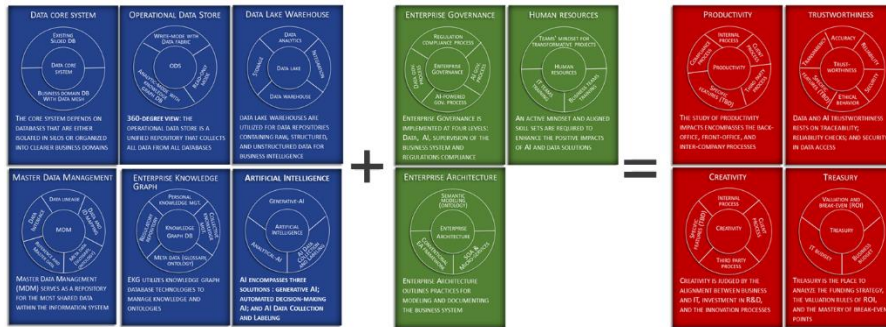
Voici le schéma complet de notre vision avec les concepts que nous venons de décrire : au centre se trouve la plateforme sémantique avec des connexions au système d'information, à l'organisation et aux algorithmes d'IA. Au-dessus, la gouvernance alimentée par l'IA (deuxième cerveau) assure le contrôle de l'ensemble du système. **Sans cette architecture, le déploiement à grande échelle de l'IA n'est ni fiable ni sécurisé. En d'autres termes, la plateforme sémantique fournit les fonctionnalités pour appliquer une « IA responsable » à travers tout le système entreprise.**





# MISE EN OEUVRE (TRAIDA FRAMEWORK)

## PRINCIPES



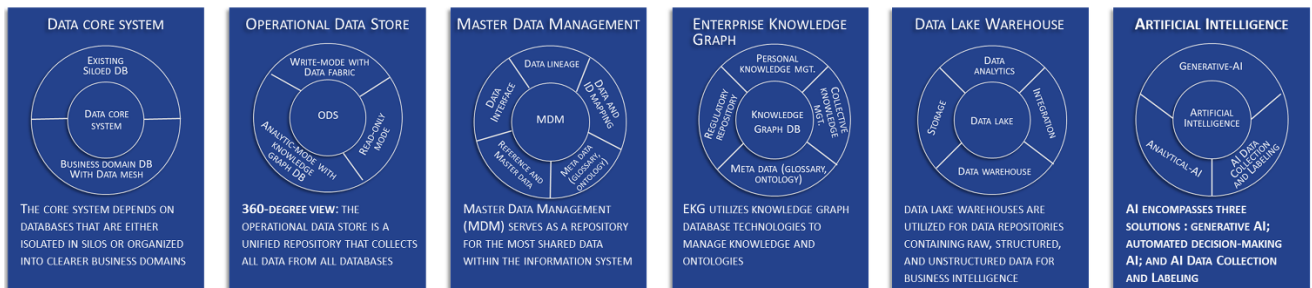
Le framework TRAIIDA (Transformative AI and Data solutions) se compose d'une série de « cartes ». Chaque carte aborde un domaine contribuant au déploiement à grande échelle des données et de l'IA. Les cartes en bleu couvrent la perspective technique, et celles en vert la perspective de gouvernance. En rouge, nous avons les

cartes du métier (business). Chaque carte permet l'accumulation de meilleures pratiques et décrit le contexte de l'entreprise. Les perspectives technique et de gouvernance doivent être alignées avec la perspective business. Cet objectif est exprimé par l'équation « IT + Gouvernance = Business ». TRAIIDA comprend 13 cartes et un total de 49 sujets divisés en 21 sujets techniques, 10 sujets de gouvernance et 18 sujets business.

Toutes ces cartes et sujets doivent être pris en compte dans le cadre de l'étude et la mise en œuvre de la plateforme sémantique pour le déploiement à grande échelle de l'IA et de ses données.

Le framework TRAIIDA aide à une analyse complète et holistique de votre contexte technique, de votre gouvernance et de l'alignement avec le business. Vous avez la possibilité de personnaliser les cartes en fonction de votre situation, d'en ajouter ou d'en supprimer, et d'adapter les sujets. Cependant, nous vous conseillons d'essayer d'abord la version initiale de TRAIIDA pour vous habituer à son utilisation.

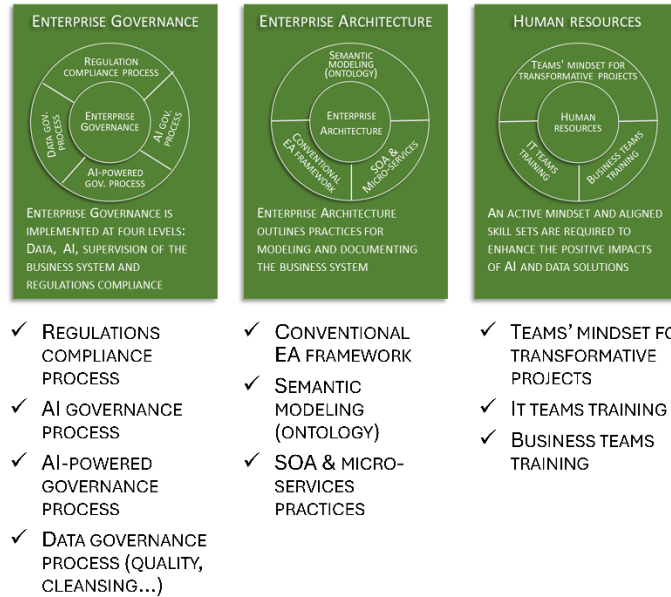
## CARTES TECHNIQUES



- ✓ EXISTING SILOED-DB
- ✓ BUSINESS DOMAIN DB WITH DATA MESH (DATA AS A PRODUCT)
- ✓ READ-ONLY MODE (INCLUDING VIRTUALIZATION)
- ✓ WRITE-MODE WITH DATA FABRIC (INCLUDING CDI)
- ✓ ANALYTICS-MODE (WITH KNOWLEDGE GRAPH DB)
- ✓ DATA LINEAGE (FLOWS)
- ✓ DATA INTERFACE (SOA)
- ✓ DATA AND ID MAPPING
- ✓ REFERENCE AND MASTER DATA
- ✓ METADATA (GLOSSARY, ONTOLOGY)
- ✓ PERSONAL KNOWLEDGE MGT.
- ✓ COLLECTIVE KNOWLEDGE MGT.
- ✓ REGULATORY REPOSITORY
- ✓ METADATA (GLOSSARY, ONTOLOGY)
- ✓ DATA ANALYTICS
- ✓ INTEGRATION
- ✓ DATA WAREHOUSE
- ✓ STORAGE
- ✓ GENERATIVE-AI
- ✓ ANALYTICAL-AI
- ✓ AI DATA COLLECTION AND LABELING

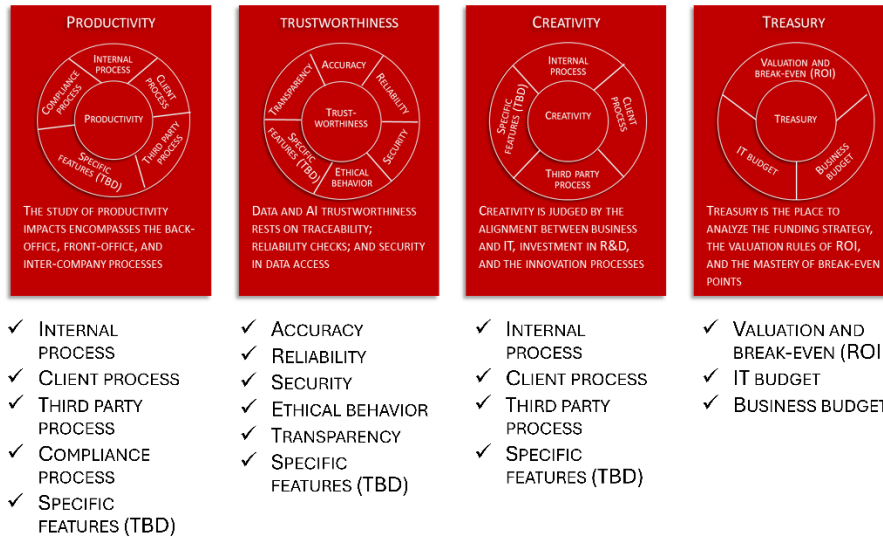
La perspective technique du framework TRAIIDA explore comment la gestion des données et les technologies de l'IA sont mises en œuvre selon les six cartes suivantes, de gauche à droite dans la figure ci-dessus : système central de gestion des données (*core system*), *operational data store* (ODS), gestion des données maîtres et de référence, base de données orientées graphes, entrepôt de données *data lake warehouse*, et l'IA. Chaque carte contient une série de sujets. Par exemple, la carte "*data core system*" inclut des sujets liés aux bases de données dans les silos et celles développées par domaine d'activité avec le *data mesh*.

## CARTES DE LA GOUVERNANCE



La figure ci-dessus montre les cartes pour la gouvernance avec leurs sujets : gouvernance d'entreprise, architecture d'entreprise et ressources humaines. Le sujet de la modélisation sémantique (ontologie) se trouve dans la carte « architecture d'entreprise ».

## CARTES DU MÉTIER



La figure ci-dessus présente les quatre cartes métier : productivité, fiabilité, créativité et trésorerie. Le framework TRAIIDA considère celles-ci comme des domaines génériques fondamentaux pour évaluer l'alignement du métier avec les perspectives techniques et de gouvernance. Pour accueillir des sujets métier dédiés à votre contexte, les trois premières cartes offrent la possibilité de définir des « fonctionnalités spécifiques » (*specific features*).

## PROCEDURE

La procédure d'utilisation du framework TRAIIDA s'exécute en trois étapes ::



1. Tout d'abord, le framework est utilisé pour évaluer votre niveau de maturité technique et de gouvernance nécessaire pour réussir vos projets de transformation avec l'IA. Les cartes en bleu et vert sont utilisées.
2. Ensuite, la perspective métier est ajoutée. Cela implique d'évaluer l'alignement entre vos objectifs métier et les niveaux de maturité technique et de gouvernance. Les cartes en rouge sont utilisées.
3. Enfin, la dernière étape se concentre sur vos projets de transformation autour de l'IA et de la gestion des données. Cela implique de construire une vision stratégique et opérationnelle, basée sur l'analyse des résultats des deux étapes précédentes.

Dans les étapes 1 et 2 de la procédure, l'analyse couvre l'ensemble du système d'information. À l'étape 3, la portée n'est plus holistique mais plutôt limitée au portefeuille de projets de transformation.

## 1. EVALUATION

Le système de notation TRAIIDA est appliqué sur les sujets de chaque carte. Il est basé sur ces trois niveaux de maturité :

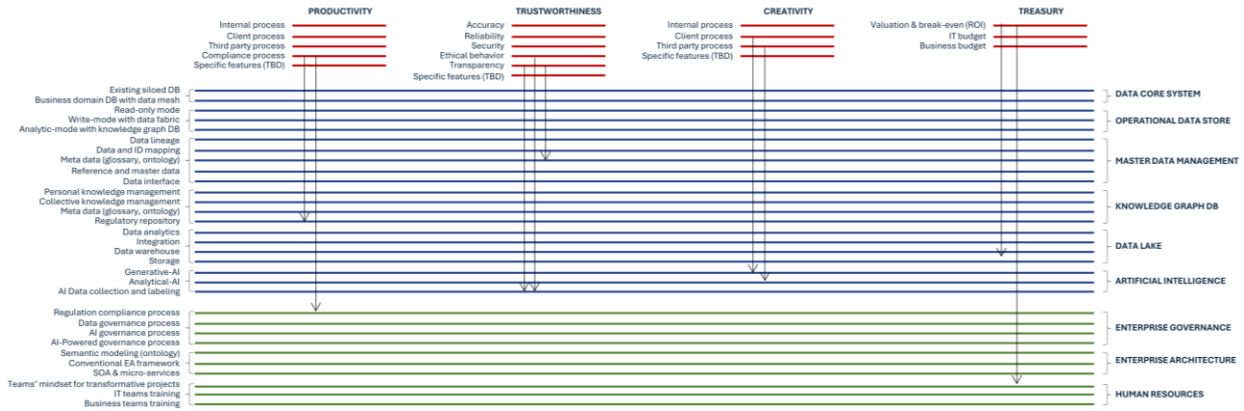
- 'Faible', indique qu'il n'y a pas d'expérience à grande échelle.
- 'Moyen', suggère que l'entreprise a commencé à explorer le déploiement à grande échelle.
- 'Élevé', signifie que l'entreprise a établi des pratiques pour le déploiement à grande échelle.

Les mesures se concentrent sur les perspectives techniques et de gouvernance, en utilisant deux formulaires mettant en évidence les scores pour le système existant et ceux pour la vision cible. Le formulaire bleu est utilisé pour les cartes techniques, et le formulaire vert est utilisé pour les cartes de gouvernance.

## 2. ALIGNEMENT AVEC LE METIER

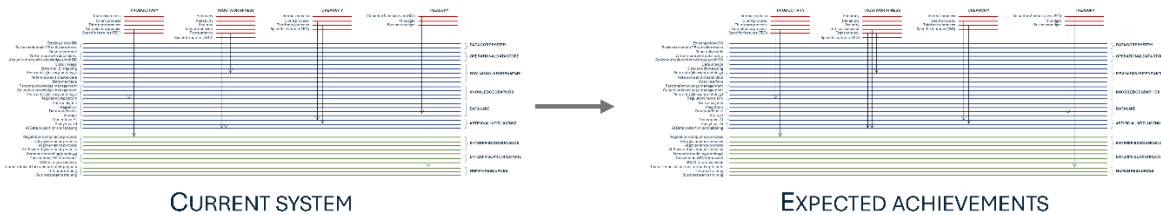
Une matrice décrit l'alignement des cartes du métier avec les cartes techniques et de la gouvernance.

Le principe consiste à tracer des lignes de dépendance des sujets du métier vers les sujets techniques et de la gouvernance. Cette cartographie concerne initialement l'analyse de la situation existante dérivée de la première étape de la procédure. Les flèches peuvent adopter le système de notation « Faible », « Moyen » et « Élevé » pour fournir plus d'informations sur la maturité de l'alignement métier avec les perspectives techniques et de la gouvernance.



Pour des raisons pratiques, il est conseillé de sélectionner des sujets métier prioritaires pour créer une première variante de la matrice. D'autres variantes s'appliquent ensuite aux sujets métier moins prioritaires.

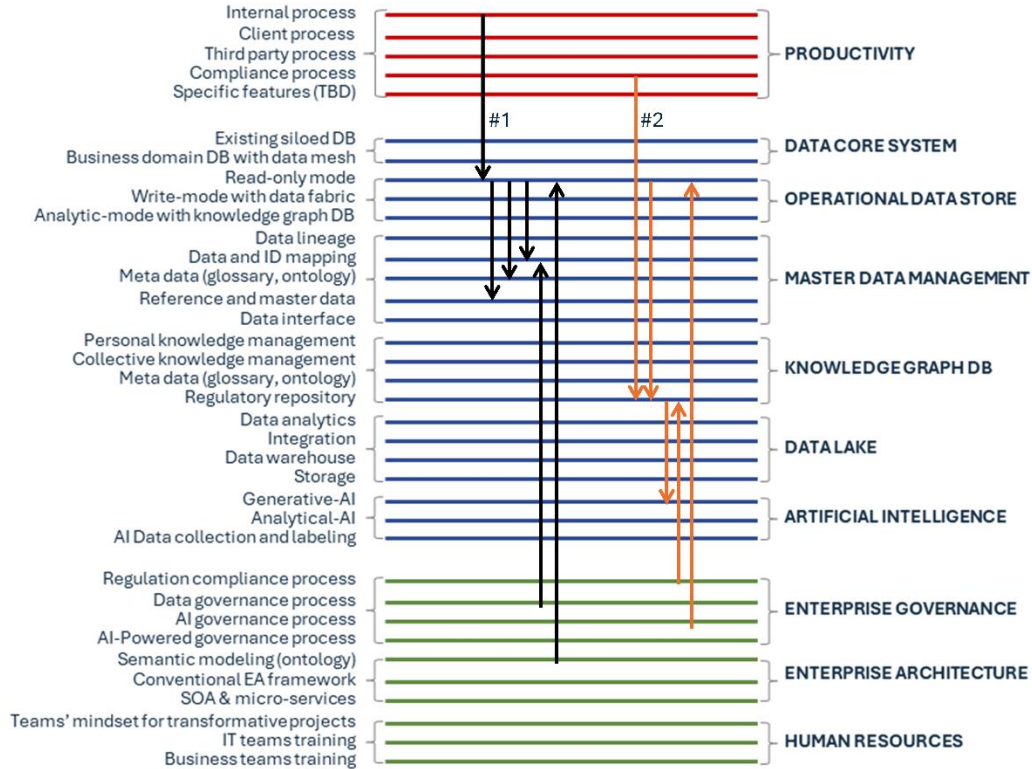
Sur la base de la matrice représentant l'état actuel du système, des objectifs d'amélioration sont identifiés pour de nouvelles perspectives techniques et de gouvernance. La matrice est ensuite mise à jour en différentes variantes pour cibler différents résultats attendus.



Grâce au framework TRAIIDA, l'étude de l'alignement métier avec la technologie est basée sur des matrices qui aident à clarifier la réflexion et la prise de décision.

### 3. PORTEFEUILLE DES PROJETS DE TRANSFORMATION

La matrice ci-dessous met en évidence un exemple de projet de transformation dans le domaine de l'intégration des données client (*Customer Data Integration – CDI*), marqué par l'objectif métier de « Productivité ».



Dans cet exemple, les flèches représentent les liens de dépendance entre les sujets du métier et ceux de la technologie et de la gouvernance. Ici, les flèches noires décrivent la mise en œuvre de la solution, et les flèches orange décrivent l'utilisation du référentiel client pour suivre la conformité réglementaire :

- #1 – Mise en œuvre de l'Intégration des Données Client – CDI (flèches noires) : Augmentation de la productivité du processus de gestion interne des clients en mettant en place un entrepôt de données opérationnelles (ODS) en lecture seule qui s'appuie sur la gestion des données de référence et maîtres (Master Data Management – MDM) et la déduplication des identifiants. La gouvernance des données est responsable de ces règles de déduplication ainsi que la gestion du cycle de vie des données. Mise en œuvre de la modélisation sémantique pour créer un modèle unifié avec une représentation à 360 degrés des clients (ontologie).
- #2 – Utilisation du CDI comme référentiel pour la conformité réglementaire (flèches orange) : copie de l'ODS dans un référentiel d'analyse de conformité réglementaire utilisant la technologie de base de données orientée graphes de connaissance. L'intelligence artificielle générative soutient la création du graphe sur les clients et aide à détecter des cas atypiques, des fraudes et des opportunités commerciales. L'ensemble du système repose sur la gouvernance d'entreprise pour la conformité réglementaire et sur la gouvernance de l'IA.

**En appliquant TRAIDA à vos projets de transformation, vous construisez et pilotez votre feuille de route. Il devient alors un excellent outil de communication pour vos parties prenantes afin d'expliquer votre stratégie en matière d'IA et de gestion des données.**

## QUELLES SONT LES PREMIÈRES ÉTAPES?

Pour saisir l'importance de la « plateforme sémantique » préconisée dans ce document, il est nécessaire d'identifier les risques associés au déploiement de systèmes d'IA sans une telle infrastructure.

Imaginez donc votre système entreprise avec de multiples applications ou agents d'IA interagissant directement avec vos bases de données, sans supervision centrale. Considérez également que plus un système augmente son intelligence, plus il devient difficile de le garder sous contrôle. Si vos systèmes d'IA produisent des résultats erronés en raison d'une mauvaise qualité des données et de failles de sécurité, vous n'aurez d'autre choix que de les déconnecter et de chercher des solutions alternatives.

Pour éviter cette situation de régression, ce rapport propose une vision d'architecture avec la « plateforme sémantique » et le framework TRAIIDA pour guider la mise en œuvre. **Cependant, une fois le contenu de ce document validé dans votre contexte, il restera encore des obstacles potentiels à surmonter avant de commencer votre parcours vers une IA fiable et responsable :**

- Ne rejetez pas systématiquement l'idée de rénover progressivement vos silos pour corriger les défauts profonds de qualité des données et rationaliser certains processus métier stratégiques ;
- Ne rejetez pas la complexité du déploiement à grande échelle de l'IA, mais utilisez-la comme un levier d'action sur le chemin de l'amélioration du système entreprise. En effet, le framework TRAIIDA peut sembler intimidant en raison du nombre de sujets qu'il couvre, mais vous contrôlez son utilisation et décidez des sujets importants. En d'autres termes, ne laissez pas la complexité entraver vos initiatives globales, mais prenez-en le contrôle. Avec l'avancement de l'IA, cet état d'esprit devient une nécessité pour en garder son contrôle ;
- L'IA est une technologie qui remplacera une partie du travail de l'humanité dans des activités intellectuelles, financières, médicales, créatives, ainsi que dans des activités physiques comme les transports, la logistique et le commerce. Pour que les individus conservent un avantage concurrentiel sur cette technologie, ils doivent cultiver un état d'esprit intéressé par leur profession pour s'améliorer continuellement et ne pas être surpassés par l'IA. En d'autres termes, les employés ne peuvent plus agir passivement dans leur travail. Ils doivent mieux comprendre leur entreprise et l'utilisation de l'IA dans leur profession. Ils doivent s'efforcer de formaliser leurs connaissances personnelles par écrit et les partager dans une approche collective de la réussite. Ils doivent développer un état d'esprit entrepreneurial au sein de leur entreprise, ce qui équivaut à une attitude « **intrapreneuriale** ». Sans une évolution de votre organisation et de vos équipes vers cet état d'esprit, il est probable que l'IA conduise à une dégradation des connaissances dans l'entreprise. Qui peut imaginer une entreprise fonctionnant uniquement avec des systèmes d'IA, sans la présence d'humains capables d'en garder le contrôle ? **Une telle organisation pourrait afficher de solides résultats pendant sa montée en puissance, mais finirait par mourir sous un effet de « consanguinité IA+IA ».** La pérennité des organisations passera par le renforcement de l'état d'esprit des équipes pour un engagement entrepreneurial avec leurs employeurs, soutenu par un couplage « Humains + IA ».

## CONCLUSION

Ce rapport est une introduction pour la prise en compte d'une vision holistique du déploiement de l'IA dans les entreprises, ainsi que des données nécessaires à sa bonne exécution. Elle est également nommée « IA responsable ». Ce sujet est difficile à aborder pour deux raisons principales :

1. **Son analyse nécessite de rassembler des expertises multidisciplinaires en IA, gouvernance des données et soutien des acteurs dans leur relation au travail.** Au sein de la communauté Engage-Meta, nous travaillons à cette union des expertises et fournissons des outils de soutien tels que le framework TRAIDA. Nos publications sont libres d'utilisation, sous une licence Creative Commons open-source.
2. Sa mise en œuvre, même si elle est progressive, peut-être une préoccupation pour les décideurs étant donné la nature transversale de l'approche. C'est une barrière habituelle pour les programmes de transformation ambitieux qui finissent trop souvent par conduire à des projets verticaux plus ou moins bien synchronisés. **Étant donné l'impact massif de l'IA sur l'avenir des entreprises, une politique molle de transversalité n'est plus viable.** C'est pourquoi ce document met l'accent sur l'importance de la vision architecturale, même si elle peut paraître conceptuelle pour les lecteurs moins expérimentés. Elle trouve une expression plus opérationnelle dans le framework TRAIDA, également décrit dans ce document. Enfin, elle est mise en œuvre concrètement avec des solutions techniques de bases de données orientées graphes de connaissances, et d'autres que nous n'avons pas énumérées dans ce rapport.

*Si vous souhaitez approfondir votre compréhension de l'architecture décrite et du framework TRAIDA, n'hésitez pas à contacter les auteurs.*





## ANNEXE : BASE DE DONNÉES ORIENTÉES GRAPHES

La technologie centrale de la plateforme sémantique repose sur les bases de données orientées graphes de connaissances. Elle permet le développement de trois types de référentiels de données nécessaires pour la plateforme :

1. La gestion des ontologies ;
2. L'accumulation des connaissances individuelles et collectives indispensables pour enrichir et entraîner les algorithmes d'intelligence artificielle.
3. Le référentiel de connaissance pour le 'second cerveau' permettant la « gouvernance alimentée par l'IA », qui inclut la création d'entrepôts de connaissances à des fins réglementaires. Par exemple, un graphe issu du texte d'une réglementation (utilisant le couplage LLM + graphe) est utilisé pour surveiller la conformité aux règles dans une industrie spécifique.

Si vous êtes novice dans la technologie des bases de données orientées graphes de connaissances, nous allons examiner les principaux critères techniques pour comprendre sa contribution. Le tableau ci-dessous inclut deux technologies pour la comparaison avec les graphes. La plus classique est l'OLTP (Online Transaction Processing) avec des outils comme Oracle ou Microsoft SQL Server, situés sur le côté droit du tableau. Au centre se trouve la technologie des graphes de connaissances avec des outils comme Neo4J, Stardog ou la plateforme IA de Palantir. Enfin, à gauche, nous présentons l'IA générative avec le LLM pour *Large Language Model* et des outils comme ChatGPT ou Claude (Anthropic). Bien que le LLM ne soit pas une base de données au sens strict, c'est néanmoins un entrepôt de connaissances. D'autres technologies existent, telles que les bases de données vectorielles ou les Data Lake Warehouse. Dans cet article, nous nous limitons à l'OLTP et au LLM, qui sont suffisants pour la comparaison avec les bases de données orientées graphes de connaissances :

	LLM	KNOWLEDGE GRAPH	OLTP
PROBABILISTIC	✓	✗	✗
DETERMINISTIC	✗	✓	✓
TRANSACTIONAL - INTEGRITY	✗	✗ ✓	✓
CARDINALITY MANAGEMENT	✗	✗ ✓	✓
HUMAN LANGUAGE READABLE	✓	✓	✗
RISK OF HALLUCINATION	✓	✗	✗
COGNITIVE CAPABILITY (E.G., INFERRED RELATION)	✓	✓	✗
UI ON STRUCTURED DATA & GOVERNANCE BUSINESS FEATURES	✗	✓	✓
DATA UPDATE ON LARGE VOLUME & REAL-TIME	✗	✓	✓
	GENERATIVE AI	SEMANTIC MANAGEMENT	TRANSACTIONAL DATA

- Seul le LLM présente un comportement probabiliste. Si vous posez plusieurs fois la même question à un outil comme ChatGPT, vous obtiendrez des réponses différentes. **Les technologies OLTP et graphes de connaissances sont déterministes.**
- Pour la gestion des transactions, c'est-à-dire l'intégrité du stockage des données, la technologie OLTP l'applique parfaitement grâce à un schéma de données formel qui décrit les structures des données et leurs contraintes. Par conséquent, l'OLTP est sécurisé mais aussi rigide lorsqu'il s'agit d'absorber des évolutions dans les structures de données et inapproprié pour les informations sans structure prédéfinie. De son côté, le LLM traite librement le texte, sans structure sous-jacente ni souci d'intégrité. **La technologie des graphes de connaissances apporte une flexibilité intéressante car elle peut fonctionner avec ou sans gestion des transactions et de l'intégrité.** En effet, les cas d'utilisation métier déterminent la mise en œuvre des règles d'intégrité. Ainsi, un concepteur de base de données en graphes de connaissances travaille sans filet par rapport à un concepteur de base de données OLTP. En d'autres termes, la flexibilité du graphe de connaissances s'accompagne du coût de la maîtrise des cas d'utilisation et des règles de modélisation associées nécessaires pour garantir l'intégrité.



- Le critère des cardinalités est au même niveau que la gestion de l'intégrité et des transactions. Une contrainte métier comme "un client ne peut être géré que par un seul commercial" est solidement mise en œuvre en OLTP et ne peut être violée. **Dans un graphe de connaissances, la mise en œuvre de cette règle est également naturelle, mais il est toujours possible de prévoir des cas d'utilisation qui vont au-delà de cette contrainte**, par exemple, si certains clients doivent être gérés par plusieurs commerciaux situés dans différentes zones. Dans le LLM, ce type de contrainte n'a pas de sens, même si l'on pourrait demander à ChatGPT de vérifier dans des documents si chaque client est attaché à un seul commercial. Il s'agit ici seulement d'une demande de contrôle de la qualité des données, et non de la création de nouvelles données.
- La capacité pour un humain de lire des données n'est pas possible en OLTP, et elle nécessite un langage technique comme le SQL. Elle a l'avantage de forcer l'utilisation d'une syntaxe et d'une grammaire formelles qui évitent les ambiguïtés. Cependant, elle est réservée à des utilisateurs techniquement qualifiés. Dans le LLM, le langage naturel est l'interface d'échange, ce qui peut induire des ambiguïtés sémantiques lorsque le texte des requêtes n'est pas bien formulé. Cependant, avec les avancées récentes dans l'analyse de données, il est désormais courant qu'une requête en langage naturel, par exemple avec ChatGPT, soit traduite en langage formel d'accès aux données comme le SQL. **Là encore, le graphe de connaissances offre une flexibilité intéressante. En effet, un humain peut découvrir et manipuler des données à travers un graphe qui ne nécessite pas de compétences informatiques.** Il est donc lisible par l'humain. De plus, divers langages techniques de création et de manipulation des données permettent une gestion formelle dans la même veine que le SQL.
- Le LLM s'accompagne du risque d'hallucination. C'est une conséquence du mécanisme de génération de texte créatif et probabiliste des LLM. **Les technologies OLTP et graphe de connaissances ne créent pas d'hallucinations.** Leur fonctionnement est déterministe.
- La capacité de raisonnement transitif existe avec le LLM et avec les graphes de connaissances. L'exemple classique dans le monde de l'intelligence artificielle est : "Tous les hommes sont mortels, Socrate est un homme, donc Socrate est mortel." **L'utilisation de cette transitivité du raisonnement est intéressante et réalisable avec la technologie des graphes de connaissances.** Elle n'existe pas en OLTP.
- La création d'applications métier est facilitée avec les technologies OLTP et graphes de connaissances. Elles offrent des cadres de gouvernance et des outils de développement d'applications rapides. Ce n'est pas encore le cas avec le LLM.
- Enfin, le dernier critère concerne la capacité des technologies à gérer de grands volumes de données en mise à jour et en temps réel. **Ici, l'OLTP se distingue, mais selon le cas d'utilisation, le graphe de connaissances peut également être utilisé.** Ce critère n'est pas pertinent pour le LLM dont le processus d'apprentissage fonctionne sur de grands volumes de données mais sans accès multi-utilisateur en temps réel. Une fois le processus d'apprentissage terminé, les utilisateurs travaillent en isolation sur leurs propres données pour la phase de personnalisation (*fine-tuning*) et ensuite interrogent avec des prompts, sans mises à jour massives de données.

-----