AURYCS
CONSULTING
STRATÉGIE DATA & IA

AURYCS INSIGHTS

Préparer ses données à l’intelligence artificielle

Pourquoi la réussite d’un projet d’IA dépend d’abord de la qualité, de l’accessibilité et de la gouvernance des données.

STRATÉGIE DATA & IA • GOUVERNANCE • AI READINESS

FONDATIONS • ACCESSIBILITÉ • CONFIANCE • IA RESPONSABLE

Une IA performante ne corrige pas des données défaillantes

Les modèles progressent rapidement. Pourtant, la qualité des réponses dépend aussi des informations auxquelles ils ont accès.

LE PARADOXE

Une organisation peut déployer un modèle avancé et obtenir des résultats peu fiables si ses données sont incomplètes, obsolètes, contradictoires ou insuffisamment protégées.

Le problème ne se limite pas à la précision du modèle. Il concerne la chaîne complète : sources, définitions, qualité, droits d’accès, transformations, indexation, recherche et contrôle des réponses.

Un meilleur modèle ne répare ni une donnée erronée, ni une autorisation mal définie, ni un système de recherche incapable de retrouver le bon document.

Trois conditions complémentaires

01 FONDATIONS

Des données comprises, qualifiées, documentées et maintenues.

02 EXPLOITABILITÉ

Une information accessible et retrouvable selon le besoin métier.

03 CONFIANCE

Des accès maîtrisés, des réponses évaluées et des usages supervisés.

De la donnée brute à une IA utile

La préparation à l’IA est une démarche d’entreprise, pas une simple succession d’outils techniques.

1. FIABILISER

Inventorier les sources, mesurer la qualité, corriger les anomalies, documenter les données et clarifier les responsabilités.

2. RENDRE EXPLOITABLE

Structurer les informations selon leurs usages : tables, documents, catalogues, moteurs de recherche et mécanismes de mise à jour.

3. GOUVERNER ET ÉVALUER

Appliquer les habilitations, protéger les données sensibles, tester la pertinence des réponses et surveiller le dispositif dans la durée.

La technologie intervient à chaque étape, mais les décisions de périmètre, de qualité et de risque restent des décisions métier et de gouvernance.

01 — Construire des fondations fiables

Avant d’entraîner, d’interroger ou de connecter une IA, il faut comprendre le patrimoine informationnel qui alimentera ses réponses.

PROFILER

Identifier les valeurs manquantes, doublons, incohérences, ruptures de format et écarts entre sources.

CORRIGER À LA SOURCE

Traiter les causes des anomalies, définir les règles de qualité et organiser les corrections récurrentes.

DOCUMENTER

Préciser la provenance, le propriétaire, la définition métier, la sensibilité et la date de dernière mise à jour.

SURVEILLER

Mesurer la fraîcheur, la complétude, la cohérence et les incidents au regard de l’usage visé.

Exemples d’outils : SQL, dbt, Great Expectations, Soda. Ils facilitent les contrôles mais ne remplacent ni les règles métier ni les responsables des données.

02 — Rendre l’information réellement retrouvable

Une donnée fiable mais introuvable au moment de la question apporte peu de valeur à un assistant IA.

DONNÉES STRUCTURÉES

Tables et référentiels : requêtes SQL, vues contrôlées, APIs, modèles sémantiques et règles d’accès. Le vectoriel n’est pas nécessaire à tous les cas d’usage.

DOCUMENTS ET CONNAISSANCES

Contrats, procédures, notes, manuels : extraction, découpage cohérent par section, métadonnées, indexation lexicale et/ou vectorielle, recherche et citation des sources.

MISE À JOUR

Synchroniser les index avec les sources, gérer les versions et les suppressions. Le CDC peut aider pour certaines bases transactionnelles ; il ne constitue pas une obligation universelle.

La bonne architecture dépend des questions à traiter, du type de données, de leur fraîcheur et des exigences de sécurité.

Exemples d’outils : LlamaIndex, LangChain, moteurs de recherche, bases vectorielles telles que Pinecone ; Kafka ou CDC selon les flux.

Comment fonctionne un assistant documentaire RAG ?

Le RAG permet à un modèle de s’appuyer sur des contenus récupérés au moment de répondre. Il ne garantit pas, à lui seul, la justesse de la réponse.

1 SOURCES

Documents approuvés, versionnés, classifiés et rattachés à un responsable.

2 PRÉPARATION

Extraction, segmentation adaptée à la structure et ajout de métadonnées utiles.

3 RECHERCHE

Index lexical et/ou vectoriel ; filtrage par droits et sélection de passages pertinents.

4 GÉNÉRATION

Le modèle construit une réponse à partir du contexte retrouvé, avec références lorsque possible.

5 CONTRÔLE

Évaluer la pertinence des sources, l’exactitude, les omissions et les réponses non étayées.

Un assistant peut produire une réponse plausible mais fausse, même lorsque les bons documents existent. La recherche et la génération doivent être testées séparément.

03 — Gouverner, protéger et évaluer

L’IA ne doit pas devenir un moyen de contourner les règles d’accès aux informations de l’organisation.

ACCÈS SELON LES HABILITATIONS

Appliquer les permissions dès la recherche, pas seulement après la génération. Prévoir les droits par rôle, document, source et contexte.

DONNÉES PERSONNELLES ET SENSIBLES

Minimiser les données, choisir la pseudonymisation ou la suppression lorsque nécessaire, protéger les index et contrôler les transferts. Retirer les identifiants avant embeddings ne suffit pas toujours à supprimer le risque.

ÉVALUATION CONTINUE

Constituer un jeu de questions métier et vérifier rappel et précision de la recherche, fidélité aux sources, exactitude, robustesse et refus appropriés.

RESPONSABILITÉS ET TRAÇABILITÉ

Désigner les propriétaires des sources, les responsables de l’usage IA, les règles de conservation et les procédures de correction ou d’escalade.

Exemples d’outils : Presidio, Ragas, DeepEval, LangSmith. Leur choix dépend du dispositif de contrôle et des contraintes de confidentialité.

Exemple — Un assistant IA pour une institution financière

Cas illustratif : une banque souhaite permettre à ses collaborateurs de retrouver rapidement des procédures internes et des règles opérationnelles.

SANS PRÉPARATION

Des procédures contradictoires coexistent, certains documents sont périmés, les droits ne sont pas homogènes et l’assistant peut citer une version obsolète.

AVEC UNE PRÉPARATION ADAPTÉE

Les procédures sont inventoriées et validées ; chaque document possède un propriétaire, une version, une date de validité et une classification. La recherche respecte les droits et cite les sources.

MESURER LE RÉSULTAT

Taux de réponses correctement étayées, pertinence des passages retrouvés, incidents d’accès, délais de recherche et satisfaction des utilisateurs.

La priorité n’est pas de connecter toutes les données de la banque, mais de sécuriser un périmètre utile, maîtrisable et mesurable.

Par où commencer ?

Une préparation progressive, pilotée par les usages et les risques, limite les investissements prématurés.

CADRER

Identifier une décision ou un processus à améliorer, ses utilisateurs, ses sources et ses critères de réussite.

DIAGNOSTIQUER

Évaluer qualité, fraîcheur, disponibilité, sensibilité, propriétaires, flux et contraintes techniques.

PRÉPARER

Corriger les données prioritaires, organiser l’accès, documenter les règles et construire un prototype contrôlé.

TESTER

Évaluer avec de vraies questions métier, vérifier les habilitations et mesurer les erreurs avant déploiement.

INDUSTRIALISER

Suivre la qualité des sources et des réponses, les coûts, les incidents et l’évolution des besoins.

La question à poser avant un projet d’IA : nos données sont-elles suffisamment fiables, accessibles et gouvernées pour cet usage précis ?

AURYCS
CONSULTING
STRATÉGIE DATA & IA

AURYCS Consulting

Préparer les organisations à l’ère de l’IA commence par leurs données.

AURYCS Consulting accompagne les organisations dans le diagnostic de leur maturité Data & IA, la gouvernance, la qualité, l’architecture des données et la définition de trajectoires d’IA responsables et créatrices de valeur.

FONDATIONS • CONFIANCE • STRATÉGIE • IA • DÉCISION • VALEUR

AURYCS INSIGHTS

Abidjan · Côte d’Ivoire | www.aurycs.com