



CHECKLIST 2 :

PRÉPARATION ET GESTION DES DONNÉES

Les données sont le carburant de l'IA générative. La qualité, la pertinence et la structure de vos données détermineront directement la performance de vos modèles. Cette checklist vous accompagne dans l'identification, la collecte, le nettoyage et la sécurisation de vos données. Elle vous aide également à mettre en place une gouvernance solide pour garantir la conformité et optimiser l'exploitation de votre patrimoine informationnel.

→ Identification et Collecte

- ☐ Recenser toutes les sources de données pertinentes, structurées et non structurées (bases de données, fichiers, cloud, API).
- ☐ Mettre en place des processus pour intégrer les données dans un référentiel centralisé et accessible.

→ Nettoyage et Qualité

- ☐ Nettoyer les données pour éliminer les erreurs, les doublons, les valeurs manquantes et les incohérences de formatage.
- ☐ S'assurer que les données d'entraînement sont de haute qualité, propres et organisées dans une structure logique.
- ☐ Valider que les jeux de données sont représentatifs, diversifiés et exempts de biais, stéréotypes ou contenu inapproprié.

→ Organisation et Structuration

- ☐ Organiser les données dans un format structuré pour faciliter leur traitement par les modèles.
- ☐ Utiliser l'IA pour classer et baliser automatiquement les documents, et enrichir les données avec des métadonnées.
- ☐ Envisager l'intégration de technologies comme la recherche vectorielle pour améliorer l'accès et la pertinence des données.

→ Sécurité et Confidentialité

- ☐ Anonymiser ou pseudonymiser systématiquement les données personnelles ou sensibles avant de les utiliser.
- ☐ S'assurer que la gestion des données est conforme aux réglementations en vigueur (ex: RGPD).
- ☐ Éviter de partager des informations confidentielles ou stratégiques avec des modèles externes non sécurisés.

→ Gouvernance des Données

- ☐ Établir une gouvernance claire pour la gestion du cycle de vie des données.
- ☐ Documenter l'origine, la transformation et l'utilisation des données pour assurer la traçabilité.
- ☐ Casser les silos de données pour permettre des flux de travail fluides et une exploitation optimale.

Des données bien préparées et correctement gouvernées sont la clé pour obtenir des résultats fiables et pertinents avec l'IA générative. En investissant du temps dans le nettoyage, la structuration et la sécurisation de vos données dès le départ, vous éviterez de nombreux problèmes en aval et maximiserez le retour sur investissement de vos projets d'IA. La qualité de vos données d'aujourd'hui détermine la performance de vos modèles de demain.

Vos données sont-elles prêtes pour l'IA générative ?

J'accompagne les organisations dans l'audit et la structuration de leur patrimoine informationnel. Mon [Diagnostic IA](#) inclut une analyse de la maturité de vos données et des recommandations concrètes pour les optimiser.

 Contact : sebastien.chami@blackbygreen.com

 Réserver un échange : <https://calendly.com/sebastien-chami/rdv>

 Formulaire de contact : <https://blackbygreen.fr/contact/>