

Rendez vos données prêtes pour l'IA, pas risquées pour l'IA

Suivez cinq étapes pour éviter les dérives de l'IA
et contourner les pièges les plus courants.



Sommaire

- 3 Pourquoi les dérives de l'IA sont plus fréquentes qu'elles ne devraient
 - 3 1. Corriger les données avant d'ajuster les modèles d'IA
 - 3 2. Mettre des garde-fous autour de l'IA
 - 4 3. Garantir la qualité et l'observabilité des données
 - 5 4. Donner à l'IA le bon contexte grâce à un graphe de connaissances
 - 5 5. Tirer le meilleur parti des outils existants
- 6 Comment Actian empêche les dérives de l'IA
- 7 À propos d'Actian



Une IA de confiance commence par des données fiables

L'IA est désormais omniprésente. Elle optimise les livraisons, aide les médecins à traiter les patients, synthétise en quelques secondes des heures d'appels clients et accélère de nombreux autres résultats.

Malgré ces réussites, beaucoup d'organisations ont aussi une histoire de « mauvaise IA » à raconter. Le problème ne se limite pas à une réponse incorrecte : les erreurs peuvent avoir des conséquences bien réelles.

- **Un chatbot fournit avec assurance une mauvaise politique de remboursement**, frustrant les clients, augmentant les appels au centre de contact et érodant la confiance.
- **Un assistant IA pourtant « intelligent » recommande un produit non pertinent**, donnant l'impression que l'entreprise ne comprend pas ses clients et les incitant à abandonner leur panier.
- **Un modèle de risque identifie les mauvais clients** tout en passant à côté de ceux nécessitant une vigilance accrue, entraînant des pertes financières, des risques de conformité et des actions correctives ultérieures.

La réaction la plus fréquente face à une mauvaise IA est souvent : « Il nous faut un meilleur modèle. » Dans la majorité des cas, la réalité est toute autre. Le problème ne vient pas du modèle.

Les résultats problématiques de l'IA trouvent presque toujours leur origine dans les données : leur gouvernance, leur surveillance et la manière dont elles sont mises à disposition des systèmes d'IA. Si elles sont incomplètes, obsolètes, biaisées ou dépourvues de contexte, même le meilleur modèle produira de mauvaises réponses, de mauvaises actions et de mauvais résultats.

Cet eBook explique pourquoi les dérives de l'IA se produisent, comment les éviter et comment bâtir une base de données fiable pour permettre à l'IA de délivrer la valeur attendue.

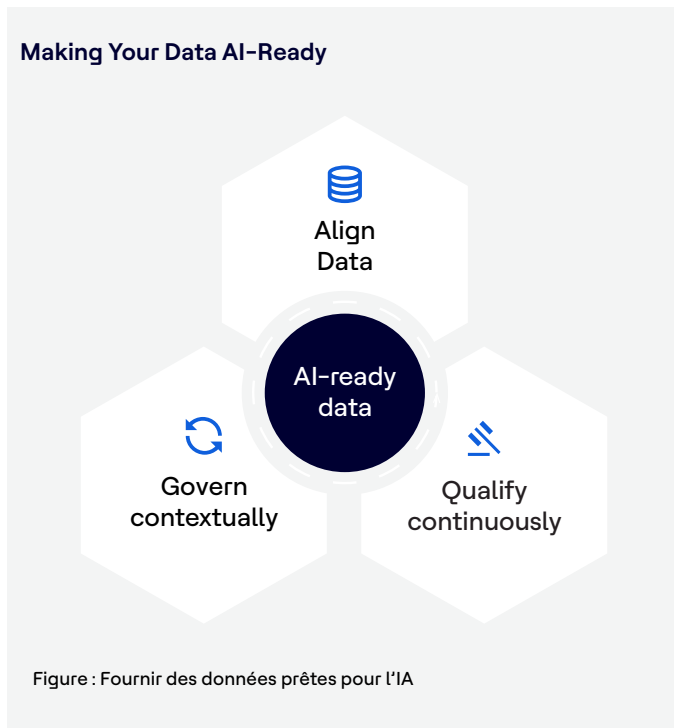


Figure : Fournir des données prêtes pour l'IA

Pourquoi les dérives de l'IA sont plus fréquentes qu'elles ne devraient

À quoi ressemble une mauvaise IA dans la pratique ? Elle prend de nombreuses formes :

- **Des réponses inexactes.** Un assistant IA « hallucine » des chiffres dans un rapport financier parce qu'il n'a pas accès aux données les plus récentes du système source.
- **Des décisions incohérentes.** Deux équipes posent la même question à un assistant de tarification et obtiennent des recommandations différentes, car chacune s'appuie sur un jeu de données distinct.
- **Des résultats biaisés ou non conformes.** Un modèle de crédit apprend à partir de données historiques biaisées et reproduit ces schémas, créant un risque réglementaire.

Dans chaque cas, le problème n'est pas que « l'IA est mauvaise ». C'est que l'IA ne disposait pas de données exactes, ne comprenait pas le contexte et n'opérait pas dans un cadre de garde-fous clairement défini.

La bonne nouvelle, c'est que ces échecs peuvent être évités. Les mêmes corrections en amont améliorent souvent les performances sur plusieurs cas d'usage, qu'il s'agisse d'un copilote de support, d'un assistant de prévision ou d'un agent d'automatisation. Comme le souligne Gartner, les données prêtes pour l'IA exigent que les organisations alignent leurs données, améliorent continuellement la qualité et gouvernent le contexte de manière adaptée (voir graphique).

Ces cinq étapes permettent de garantir des résultats IA solides et d'éviter les pièges courants, aidant les organisations à passer de dérives ponctuelles à des résultats fiables et reproductibles.

1. Corriger les données avant d'ajuster les modèles d'IA

Il est tentant de se concentrer sur les modèles, les prompts et les nouveaux outils d'IA, car ce sont les éléments les plus visibles. Pourtant, le vrai travail se situe en coulisses, au niveau des fondations de données.

Si un système d'IA répète des erreurs évidentes, peine à répondre à des questions simples ou fournit des réponses auxquelles personne ne fait confiance, la cause racine est très souvent liée aux données. Le problème peut concerner :

- **La qualité des données.** Les informations sont-elles exactes, complètes et à jour ?
- **Le contexte.** L'IA comprend-elle comment les actifs de données sont liés entre eux ?
- **La gouvernance.** Des règles encadrent-elles réellement l'usage des données ?

Avant d'investir davantage de temps et d'argent dans les modèles ou les outils, les organisations doivent prendre du recul et corriger les données. Lorsque la qualité, le contexte et la gouvernance sont en place, l'IA s'appuie moins sur des suppositions et davantage sur des signaux validés. Elle fournit des réponses auxquelles les équipes font confiance, des décisions qu'elles peuvent défendre et des résultats qui font avancer le métier.

2. Mettre des garde-fous autour de l'IA

Lorsque l'IA prend des décisions ou génère du contenu, les dirigeants s'inquiètent des risques. Les questions reviennent souvent :

- « Et si l'IA exposait des données sensibles ? »
- « Et si elle formulait des recommandations contraires aux politiques internes ? »
- « Et si un régulateur demandait comment une décision a été prise et que nous étions incapables de l'expliquer ? »

Ces préoccupations mettent en évidence le rôle central de la gouvernance des données. La gouvernance regroupe l'ensemble des règles, processus et responsabilités qui définissent quelles données existent, qui en est responsable, qui peut les utiliser, dans quel but, et comment la qualité et les accès sont contrôlés.

On peut l'assimiler à un système de circulation. Sans feux tricolores, voies de circulation et limitations de vitesse, les routes deviennent dangereuses et imprévisibles, en particulier lorsque des agents agissent de manière autonome ou que les décisions impactent les clients, les finances ou la conformité.

Voici pourquoi la gouvernance des données est essentielle pour l'IA :

- **Explicabilité.** Lorsqu'une IA recommande de refuser un prêt ou d'approuver une demande, il faut pouvoir retracer les données utilisées et comprendre pourquoi.
- **Conformité.** Des réglementations comme le RGPD ou des règles sectorielles imposent une visibilité sur l'usage des données sensibles.
- **Équité.** La gouvernance aide à détecter des données d'entraînement biaisées et à éviter qu'elles n'influencent les décisions de l'IA.

Concrètement, la gouvernance devient effective lorsque les équipes désignent des responsables de données clairs, définissent les usages autorisés de l'IA, mettent en place des règles d'accès et assurent le suivi du lignage. Ainsi, la qualité et la responsabilité ne disparaissent pas lorsque l'on passe d'un pilote à la production.

Les termes clés de la donnée pour l'IA, expliqués simplement

- **Agent d'IA.** Logiciel qui exécute des actions à partir de données et de règles. Il peut, par exemple, lire des tickets de support, formuler des réponses et interagir avec des clients.
- **Lignage des données.** Traçabilité de l'origine d'un jeu de données, de ses transformations dans le temps et de ses usages.
- **Produit de données.** Jeu de données structuré, réutilisable, adossé à un contrat de données et conçu pour être utilisé en toute sécurité par des systèmes d'IA et des utilisateurs.
- **Gouvernance des données.** Ensemble des politiques et processus qui encadrent la création, l'accès et l'usage des données.

3. Garantir la qualité et l'observabilité des données

Même avec une bonne gouvernance, l'IA peut dériver. Cela se produit lorsque les données alimentant les modèles sont inexacts, incomplètes ou sujettes à une dérive, c'est-à-dire lorsque les comportements et les schémas sous-jacents évoluent sans que le modèle ne l'ait appris.

Voici comment cela peut conduire à une véritable dérive de l'IA. Imaginez un distributeur utilisant une IA fondée sur les données de ventes, les niveaux de stock et les promotions. Du jour au lendemain, un pipeline de données se rompt et les ventes de la veille pour une région donnée n'arrivent jamais.

L'IA ne sait pas que les données sont manquantes. Elle interprète cette absence comme une chute de la demande et recommande de réduire les commandes. Lorsque les équipes constatent des rayons vides, l'erreur de l'IA est déjà devenue un problème métier concret.

Pour éviter ce scénario, les organisations ont besoin de capacités de qualité et d'observabilité des données. Ces mécanismes valident les données critiques et détectent les problèmes en amont, avant qu'ils ne soient amplifiés par les systèmes d'IA.

Les étapes clés sont les suivantes :

- **Mettre en place des contrôles simples**, en commençant par des tests de base sur les données les plus critiques.
- **Surveiller la dérive des données**, qui survient lorsque leur signification évolue dans le temps, par exemple avec un nouveau profil de clients, de nouveaux codes produits ou une nouvelle manière d'enregistrer les résultats.
- **Boucler la résolution**, en s'assurant qu'un problème détecté est rapidement attribué au bon responsable et suivi jusqu'à sa correction.

5 signaux d'alerte précoces d'une mauvaise IA

- Les équipes métier contournent régulièrement les recommandations de l'IA parce qu'« elles ne semblent pas correctes ».
- Des rapports ou tableaux de bord différents donnent des réponses différentes à la même question.
- Les problèmes sont découverts suite à des plaintes clients, et non via des alertes internes.
- Personne ne sait précisément quelles sources de données alimentent un cas d'usage IA donné.
- Les nouveaux projets d'IA stagnent au stade de la préparation des données.

Chacun de ces signaux indique que la fiabilité des données et la gouvernance doivent être renforcées avant de passer à l'échelle.

4. Donner à l'IA le bon contexte grâce à un graphe de connaissances

L'IA n'a pas seulement besoin de faits. Elle a aussi besoin de contexte : définitions partagées, relations et signification métier. Par exemple, une organisation peut vouloir répondre à une question apparemment simple : « Montre-moi nos clients à forte valeur qui ont eu une interaction négative avec le service le mois dernier et qui présentent un risque de churn. »

Pour y répondre correctement, l'IA doit comprendre :

- **Qui** est considéré comme un « client »
- **Ce** qui définit un client « à forte valeur »
- **Quelles** interactions sont jugées négatives
- **Comment** le risque de churn est défini dans l'entreprise et où se trouve cette définition

Dans de nombreuses organisations, ces définitions sont réparties dans différents systèmes, détenus par différentes équipes et interprétés de diverses manières. Si l'IA ne perçoit pas les relations ou ne sait pas quelle définition fait autorité, elle devine... et échoue le plus souvent.

Un graphe de connaissances permet de résoudre ce problème. Il agit comme une cartographie de l'entreprise, capturant des entités clés telles que les clients, commandes, produits ou régions, ainsi que les relations entre elles.

Lorsque l'IA dispose de ce contexte structuré, elle peut répondre avec davantage de précision. Le contexte lui permet aussi de comprendre comment une modification à un endroit donné impacte les usages et dépendances en aval.

Pour garantir ce contexte :

- **Commencer par un actif de données à fort impact**, lié par exemple à l'expérience client, à la chaîne d'approvisionnement ou à la performance financière.
- **Formaliser un langage commun** en définissant clairement des notions comme « client prioritaire », « incident ouvert » ou « compte à risque ».
- **Relier métadonnées techniques et signification métier** dans le catalogue de données afin que l'IA et les utilisateurs puissent les interpréter correctement.

5. Tirer le meilleur parti des outils existants

Beaucoup d'organisations pensent devoir adopter de nouveaux outils pour réussir avec l'IA. Bien souvent, c'est l'inverse. La plupart disposent déjà d'un empilement technologique dense incluant plateformes de données, catalogues, pipelines, outils de BI et solutions de machine learning, avec des fonctionnalités qui se chevauchent mais sont peu adoptées ou mal exploitées.

De la même manière, les pilotes d'IA émergent souvent en parallèle. Différentes équipes développent des assistants ou des agents à partir de jeux de données, de définitions et de règles distinctes. Le résultat est prévisible : des réponses contradictoires, des efforts dupliqués et une IA difficile à faire évoluer, à gouverner et à fiabiliser.

La véritable opportunité n'est pas d'ajouter un outil de plus, mais de faire fonctionner ensemble les outils et les données existants autour de standards partagés, de produits de données communs et de garde-fous réutilisables pour chaque cas d'usage IA.

Parmi les actions concrètes :

- **Cartographier le patrimoine de données** en recensant les plateformes, catalogues, pipelines et outils d'IA, afin d'identifier les doublons et les silos.
- **Standardiser des produits de données réutilisables**, plutôt que de laisser chaque équipe recréer ses propres jeux de données.
- **Mesurer l'adoption**, et pas seulement la disponibilité, en suivant l'usage réel des produits de données et des assistants IA.

Feuille de route en 7 étapes vers de meilleurs résultats IA

1. **Choisir un cas d'usage IA précis**, comme un assistant commercial, un agent de tri du support client ou un agent de prévision de la chaîne d'approvisionnement.
2. **Tracer les données sous-jacentes au cas d'usage** en identifiant les jeux de données clés, leurs responsables, les contrôles qualité et les règles d'accès.
3. **Garantir la qualité et l'observabilité** en ajoutant des tests et des mécanismes de surveillance.
4. **Clarifier la gouvernance et les garde-fous** en documentant les données accessibles et les résultats acceptables.
5. **Ajouter le contexte** via un graphe de connaissances ou une couche sémantique.
6. **Piloter avec de vrais utilisateurs**, en maintenant l'humain dans la boucle, capables de comprendre à la fois les données et la logique métier. Identifier les dérives de l'IA et les transformer en améliorations concrètes des données, des règles ou du contexte.
7. **Formaliser ce qui fonctionne**. Transformer les enseignements tirés en modèles réutilisables et les appliquer au cas d'usage IA suivant.

Comment Actian empêche les dérives de l'IA

Actian rend les données découvrables, fiables et exploitables afin que l'IA puisse tenir ses promesses. La plateforme Actian Data Intelligence soutient une « bonne IA » en fournissant :

- **Un raisonnement IA précis**, grâce à un graphe de connaissances fédéré qui apporte aux agents d'IA un contexte métier, des définitions et des relations cohérentes entre les domaines.
- **Des entrées validées en temps réel pour l'IA**, via un serveur Model Context Protocol (MCP) qui alimente les agents et les outils analytiques avec exactement les données dont ils ont besoin, dans le format attendu.
- **Des données contextualisées par leur origine**, grâce au lignage des données, qui retrace précisément d'où proviennent les données et comment elles évoluent, offrant à l'IA la clarté nécessaire pour éviter les approximations et produire des résultats fiables.
- **Des données fiables pour les modèles d'IA**, soutenues par la qualité des données et l'observabilité, qui surveillent les pipelines, détectent les anomalies et traitent les problèmes avant qu'ils ne se transforment en dérives de l'IA.
- **Une gouvernance intégrée dès la conception**, en intégrant politiques, lignage et contrôles d'accès dans les workflows quotidiens, permettant aux domaines de garder le contrôle de leurs données sans freiner l'innovation.
- **Des résultats IA plus rapides et plus fiables**, grâce à des produits de données soutenus par des contrats de données garantissant la livraison, la gouvernance et l'intégrité à chaque point d'interaction.

Plutôt que d'assembler une multitude d'outils disparates, les organisations bénéficient d'une base de données unifiée, conçue pour une IA digne de confiance. La plateforme Actian résout des problématiques de données qui vont bien au-delà d'un simple projet IA isolé.

Les organisations gagnent ainsi une confiance accrue dans leurs données, réduisent le délai entre l'idée et l'impact réel de l'IA, et limitent le risque d'erreurs coûteuses liées à l'IA. En investissant dès aujourd'hui dans la qualité, le contexte et la gouvernance, elles créent un environnement dans lequel les agents d'IA et les utilisateurs métier peuvent agir en toute confiance, et où les équipes font confiance aux réponses et aux actions produites.

C'est ainsi que les organisations passent de dérives ponctuelles de l'IA à des résultats durables. Elles optimisent et déploient une IA précise, explicable et alignée sur les résultats qui comptent le plus pour le métier.

À propos d'Actian

Actian aide les entreprises à gérer et gouverner leurs données en toute confiance, à grande échelle. Les organisations font confiance aux solutions de gestion et d'intelligence des données d'Actian pour maîtriser des environnements data complexes et accélérer la mise à disposition de données prêtes pour l'IA. Conçues pour être flexibles et fiables, ces solutions s'intègrent facilement à tous types d'environnements : sur site, dans le cloud ou en mode hybride. Actian est la division Data & IA de **HCLSoftware**. Pour en savoir plus : www.actian.com.