



# Intelligence artificielle

## Guide d'une IA décomplexée pour votre entreprise

RETOURS D'EXPÉRIENCE DE NOS CONSULTANTS MERITIS

**Charbel El Fakhry** • **Antony Arslanyan** • **Abdelkader Hamadi** • **Amady Ba**



<b>Bio de l'auteur</b> _____	<b>PAGE 03</b>
<b>Introduction</b> _____	<b>PAGE 04</b>
<b>Partie 1 • Comment créer un modèle IA efficient ?</b> _____	<b>PAGE 05</b>
<b>#1</b> Pourquoi intégrer un modèle intelligent ? _____	<b>PAGE 06</b>
<b>#2</b> Quels types de données ai-je en ma possession ? _____	<b>PAGE 08</b>
<b>#3</b> Quel type d'IA mettre en place ? _____	<b>PAGE 10</b>
<b>#4</b> Quelle application et quel algorithme associer à quel type d'apprentissage ? _____	<b>PAGE 12</b>
<b>Partie 2 • Comment implémenter son modèle?</b> _____	<b>PAGE 14</b>
<b>#1</b> Préparer ses données _____	<b>PAGE 15</b>
<b>#2</b> Trouver quel type de modèle tester _____	<b>PAGE 17</b>
<b>#3</b> Le prototype _____	<b>PAGE 19</b>
<b>Conclusion</b> _____	<b>PAGE 21</b>

## [Bio des auteurs]

### Charbel EL FAKHRY • Data Scientist

Charbel est un spécialiste d'Intelligence Artificielle et Data science.

Il est diplômé d'un Master en IA de l'UPMC spécialisé en données, apprentissage et connaissances.

Il dispose également d'un diplôme d'ingénieur en Informatique de la Lebanese American University.

Charbel a travaillé sur de nombreux projets innovants en Machine / Deep Learning et Intelligence Artificielle, dont un système de reconnaissance d'émotions grâce à la reconnaissance vocale et faciale, la reconnaissance des émotions de la parole et la classification de l'âge et du genre.

Il intervient sur divers projets d'IA innovants chez des grands clients principalement dans le secteur financier tels que le Crédit Agricole et la Société Générale.



### Antony ARSLANYAN • Ingénieur Intelligence Artificielle

Antony est Data Scientist spécialisé dans l'entraînement de modèle Deep Learning.

Au cours de ses différents projets il a acquis une expertise dans le domaine du NLP (traitement du langage) principalement en utilisant Bert et HuggingFace, mais également de l'expérience en Computer Vision.

Passionné, Antony s'intéresse à tout ce qui touche à l'Intelligence Artificielle et plus particulièrement à la technologie des Transformers.



### Abdelkader HAMADI • Ingénieur Chercheur en Data-Science

Abdelkader est spécialisé dans les domaines de l'intelligence artificielle, la recherche d'information, la vision par ordinateur et de la data. Ingénieur, puis Docteur en Informatique depuis 2014 ; il a rejoint Meritis 2019 en tant qu'ingénieur-conseil en Recherche et Développement.

Actuellement, il collabore avec la SNCF sur un projet innovant visant à développer des solutions de surveillance automatique du réseau ferroviaire par la fibre optique. Précédemment, il a occupé un poste d'Attaché Temporaire d'Enseignement et de Recherche à l'université de Lorraine de 2013 à 2015. Puis a assuré un poste de Maître de Conférence à l'université de Mostaganem (Algérie) de 2015 à 2019.



### Amady BA • Data Scientist et ingénieur Machine Learning

Féru de la Data, Amady est titulaire d'un Master en Ingénierie Mathématique option Data Science. Par la suite, il a eu une première expérience de 3 ans dans une entreprise spécialisée en assurance santé en tant que Data Scientist.

Sa volonté de relever de nouveaux défis est en effervescence permanente. Ainsi en 2021, il rejoint Meritis en tant que Consultant en Data Science.

Aujourd'hui, Amady accompagne nos clients à mettre en place des projets R&D allant de la conception jusqu'à l'industrialisation.





# Introduction

**Le 8 novembre 2021, le Gouvernement lançait la deuxième phase de sa stratégie nationale pour l'IA. Objectif : accroître le nombre de talents formés dans ce domaine, et transformer le potentiel de recherche et développement en succès économiques. Une ambition à laquelle l'État prévoit de consacrer au total 2,22 milliards d'euros sur cinq ans.**

Pourquoi un tel plan en faveur de l'intelligence artificielle ? Parce que le constat est aujourd'hui sans appel : 85 % des projets d'IA échouent et aboutissent à des résultats erronés en raison de préjugés dans les données, des algorithmes ou des équipes responsables. Résultat, seuls 15 % des cas d'utilisation tirant parti des techniques d'IA impliquant des environnements Edge et IoT parviennent à passer en production.

Une proportion alarmante mais néanmoins étonnante quand on sait qu'en 2021, 71 % des entreprises exploraient ou exploitaient déjà le potentiel de l'IA ! L'IA n'est aujourd'hui plus une chimère ou un fantasme hollywoodien, mais bel et bien une technologie de plus en plus mature et surtout un levier de création de valeur hors norme... à condition de réussir sa mise à l'échelle. En effet, passer du PoC à la phase d'industrialisation implique de maîtriser les algorithmes et surtout de ne pas brûler les étapes.

C'est là que les entreprises bloquent le plus souvent, faute de ressources et de moyens en interne, mais aussi de « savoir faire ». Par où commencer ? Quels prérequis pour réussir son passage à l'IA ? Quelles sont les best practices et les erreurs à éviter ? Comment faire de l'IA un facteur d'aide à la décision stratégique ? Autant de questions auxquelles elles ont souvent du mal à répondre et qui affectent les performances de leurs modèles intelligents.

Sur la base des retours d'expérience terrain de nos experts, découvrez toutes les étapes pour réussir la mise en place de modèles efficaces jusqu'à leur mise en production.

[Partie 1]

# Comment créer un modèle IA efficient ?

## Pourquoi intégrer un modèle intelligent ?

Selon les prévisions d'IDC, le volume de données mondiales devrait dépasser 180 zettaoctets à l'horizon 2025, soit une croissance annuelle moyenne de près de 40 % sur cinq ans. Une masse exponentielle impossible à exploiter sans l'aide de l'IA.

### Les avantages majeurs de l'IA

#### 1 Apporter une aide à la prise de décision stratégique

Comment tirer le meilleur parti de ses données lorsque l'on doit composer avec des milliers voire des millions de données ? Voilà l'atout n°1 de l'IA : faciliter et accélérer les prises de décision en triant les données selon leur pertinence et en apportant rapidement aux décideurs une information qualifiée.

#### 2 Exploiter les données de façon optimale

Collecter les données, c'est bien, mais savoir les exploiter, c'est mieux. Or impossible désormais de traiter les Téraoctets de data qui transitent sur l'ensemble des canaux de l'entreprise (réseaux sociaux, objets connectés, service client, site e-commerce...). Le risque pour l'organisation ? Passer à côté d'informations stratégiques. Un agent intelligent permettra alors d'identifier de nouvelles pistes d'exploration auxquelles n'auront pas accès vos concurrents.

#### 3 Automatiser les process

In fine, l'intelligence artificielle permet surtout d'automatiser les tâches chronophages et à faible valeur ajoutée. À la clé : un gain de temps précieux et la possibilité pour les équipes de se concentrer sur leur cœur métier. L'IA permet ainsi de réaliser certaines tâches jusqu'alors impossibles à réaliser manuellement.

### Les principaux cas d'usage de l'IA

La liste est loin d'être exhaustive, mais certains secteurs sont déjà bien avancés dans l'utilisation de l'IA. Par exemple :

- La finance avec la détection des fraudes ou la prédiction du stock market pour les jours à venir ;
- La médecine pour trouver des anomalies dans les molécules, détecter des facteurs à risque dans certaines maladies...
- Le retail via la recommandation automatique : suggestion de visites à effectuer ou de contenus à visionner, services associés à un achat, produits complémentaires...

## [Question #1]

### » L'avis de l'expert

Nous faisons face aujourd'hui à de très gros volumes de données. Pour effectuer des analyses prédictives, les entreprises n'ont désormais plus le choix et doivent recourir à des modèles IA, notamment dans des campagnes de prévention santé. Pour accompagner au mieux leurs adhérents, des mutuelles anticipent ainsi leurs potentiels problèmes de santé par le biais de campagnes de prévention faisant appel à des modèles IA. À la clé : la capacité à détecter les pathologies à risque pour l'adhérent et lui proposer un accompagnement adapté.



**Amady Ba**  
Consultant  
Data Scientist  
& Machine  
Learning  
Engineer

Les entreprises les plus matures en IA ont vu leur chiffre d'affaires augmenter jusqu'à

# 50%

**par rapport  
à celui de leurs pairs.**

Source : [Rapport de recherche Accenture, juin 2022](#)

## Quels types de données ai-je en ma possession ?

Avant de mettre en place un modèle intelligent, la première étape consiste à dresser un état des lieux précis des différents types de données disponibles afin d'opter par la suite pour le modèle le plus adapté.

### Les différents types de données

#### 1 Les données structurées

Il s'agit des données disponibles dans les bases de données relationnelles, bien organisées sous forme de ligne de table (par exemple : nom, prénom, âge...).

#### 2 Les données semi-structurées

Ces données font écho à un format particulier de type clé-valeur le plus souvent utilisé pour envoyer des messages entre serveurs et clients.

#### 3 Les données non structurées

De quoi s'agit-il ? Tout simplement de tous les autres types de format de données : images, vidéos, PDF, documents, e-mails, etc.

### Les étapes de traitement des données

Que souhaitez-vous faire de vos données ? Quel est l'objectif souhaité ? Quel est le cas d'usage associé ? Vos données sont-elles étiquetées ou non ? Pour mettre en place un modèle d'intelligence artificielle efficient, vous devez au préalable décider quelles données utiliser et dans quel contexte.

Ensuite, suivez le guide pour parvenir à des données de qualité :

- D'abord, il importe de comprendre le métier autant que vos données : si vous souhaitez construire un modèle, vous devez d'abord connaître le domaine concerné et son contexte. Dans un environnement médical par exemple, seul un expert sera capable de dire quelle donnée est fautive, incomplète...
- Puis, vous pouvez collecter, nettoyer et enrichir les données.
- Enfin, vous pouvez passer à la construction des indicateurs (ou des features) selon les différents scénarios établis.

**Un travail souvent négligé alors qu'il représente 80 % de la performance d'un modèle. De la qualité de vos données découle l'efficacité de votre agent intelligent. La préparation et l'exploration des données constituent ainsi la première phase de construction de votre modèle.**



## [Question #2]

### ” L’avis de l’expert

Je travaille principalement sur des données non-structurées souvent difficiles à aborder. C’est pourquoi, avant de commencer le travail de structuration, il est nécessaire de prendre le temps d’analyser les données pour éviter de créer des biais dans le modèle. Dans le cadre d’une mission chez un assureur, nous avons créé un agent intelligent afin d’automatiser le processus d’intégration des informations issues des documents d’identité de leurs adhérents. La première étape consistait à reconnaître le type de document dont il s’agissait : passeport, permis de conduire, carte d’identité... Or nous n’avions entraîné le modèle que sur des papiers français. Par conséquent, les documents étrangers n’étaient pas reconnus. Dans ce processus, il est important de savoir que d’autres types de sources de données existent même si on ne les traitera jamais.



Antony  
Arslanyan  
Ingénieur  
Intelligence  
Artificielle

Les lacunes dans la gestion  
des données empêchent

# 78%

**des entreprises  
de tirer pleinement parti  
de leurs données.**

Source : [“State of Enterprise CRM Data Management”](#),  
réalisé par Forrester pour Odaseva, 2021

## Quel type d'IA mettre en place ?

Maintenant que vous avez une vision claire sur la nature de vos données, que souhaitez-vous que la machine prédise ? La question peut paraître simple, mais dont la réponse aura un impact direct sur le mode d'apprentissage retenu.

### Quel type d'apprentissage choisir ?

Il existe plusieurs types d'apprentissages spécifiques, mais deux d'entre eux sont le plus souvent utilisés par les data scientists.

#### 1 L'apprentissage supervisé

Cette méthode s'emploie avec des données étiquetées, c'est-à-dire qui contiennent déjà une valeur observée. Ici, vous connaissez précisément le type d'information ou de résultat que vous voulez obtenir en sortie. Prenez alors le temps de labelliser correctement les données, c'est ce qui vous permettra d'avoir des résultats cohérents, et de tester ensuite.

#### 2 L'apprentissage non supervisé

Autre option : le mode non supervisé lorsque vous ne savez pas ce que vous souhaitez faire de vos données, ni comment les exploiter. Des données qui, ici, ne sont pas étiquetées. L'idée est alors de construire un modèle reposant sur des habitudes de comportement pour détecter les anomalies ou recommander un contenu similaire.

### Les étapes pour choisir le bon type d'apprentissage

Attention de ne pas choisir le modèle avant les données ! Vous devez au contraire définir le mode d'apprentissage le plus adapté selon les données vous avez à disposition. Voici les questions à vous poser pour pouvoir créer le modèle adéquat :

- Quelle tâche souhaitez-vous que le modèle intelligent exécute ?
- Quelle architecture de modèle sera la plus fonctionnelle ?
- Souhaitez-vous faire du deep learning ou du machine learning ? (Voir section suivante pour plus de précisions sur ce point).

Pour vous aider dans vos choix, il est important de mener une veille active sur les techniques déjà existantes, de lire des articles, de vous renseigner sur ce qui se fait dans d'autres langues... pour savoir comment ça va marcher et s'il sera facile de faire ingérer la donnée par le modèle.

## [Question #2]

### » L'avis de l'expert

Il existe un modèle intermédiaire qui repose sur des cas semi-supervisés pour lesquels seule une partie des données est étiquetée. Prenons l'exemple d'un détecteur de navette spatiale : vous devez disposer de suffisamment de photos pour la différencier d'un avion. Ici, seule une partie des données est étiquetée et on risque très certainement de se heurter à un problème de manque de données. En effet, une méthode supervisée imposerait de classifier les images étiquetées « non-navette spatiale » et les photos de « navette spatiale » pour lesquelles peu de données sont disponibles. Certaines méthodes dites semi-supervisée existent alors pour contourner la problématique, comme la méthode SVM 1-class qui permet d'apprendre malgré le manque de données.



**Abdelkader Hamadi**  
Ingénieur  
Chercheur en  
Data Science

# 1 entreprise 2

**est dans l'incapacité d'exploiter ses propres données clients.**

Source : [Rapport "State of Enterprise CRM Data Management"](#), réalisé par Forrester pour Odaseva, 2021

## Quelle application et quel algorithme associer à quel type d'apprentissage ?

Un seul algorithme ne peut pas répondre à toutes les problématiques. Quel modèle algorithmique alors mettre en œuvre selon l'application souhaitée, le type de données et les objectifs souhaités ?

### Des modèles liés à l'apprentissage non supervisé

#### 1 Le clustering

L'apprentissage non supervisé est principalement utilisé en matière de clustering (également appelé clusterisation ou association). L'idée est d'associer au sein d'un même dataset des données qui se ressemblent en regroupant un ensemble d'éléments hétérogènes sous forme de sous-groupes homogènes ou aux caractéristiques similaires.

##### ➤ Algorithmes

K-Means / Isolation Forest / DB Scan / Local Outlier Factor (LOF)

#### 2 La réduction de dimensions

L'apprentissage non supervisé peut aussi permettre d'améliorer les performances de l'apprentissage supervisé dans le cas d'un grand nombre de paramètres en entrée. La technique du regroupement ou de réduction est alors utilisée pour réduire le nombre total de dimensions.

### Les modèles liés à l'apprentissage supervisé

#### 1 La classification

Ce problème se pose quand on a au moins deux voire plusieurs catégories en sortie. Par exemple : reconnaître si l'image est celle d'un chat ou d'un chien, si une donnée représente une personne malade ou en bonne santé, etc.

On peut également y associer la notion de segmentation qui consiste à identifier tous les pixels associés à l'objet de la recherche dans une image. On classe alors par rapport aux pixels et non à l'image entière.

##### ➤ Algorithmes

Régression logistique / Support Vector Machine (SVM) / Arbre de décision / Random Forest / GBM / Naïve Bayes

#### 2 La régression

Ce problème permet de prédire une valeur continue. Par exemple : quel est le prix d'une maison. La régression linéaire, très simple à mettre en place, est souvent le premier algorithme que l'on apprend quand on manipule du machine learning.

##### ➤ Algorithmes

Régression linéaire / Arbre de décision / Random Forest / GLM : modèle linéaire généralisé / Gradient Boosting Machine (GBM)



## [Question #4]

### » L'avis de l'expert

Le choix d'un algorithme dépend de la problématique et du type de données, et impose de bien connaître l'état de l'art. Certains modèles algorithmiques fonctionnent très bien pour certaines problématiques données. Cette connaissance peut vous faire gagner beaucoup de temps. Essayez donc toujours d'observer la performance de vos modèles sur des données similaires ou à partir de problématiques semblables. Si votre modèle n'a jamais été traité, élaborer des hypothèses et testez des modèles différents pour voir quelle catégorie donne la meilleure performance.



Abdelkader  
Hamadi  
Ingénieur  
Chercheur en  
Data Science

Seulement  
**50%**

**des décideurs déclarent que leur organisation est engagée dans une démarche de valorisation des données.**

Source : [Étude OpenDataSoft, décembre 2021](#)



**[Partie 2]**

# **Comment implémenter son modèle ?**

## Préparer ses données

Maintenant que vous avez une vision 360 de vos données, il est temps de préparer vos données. Comment ? Tout dépend des objectifs liés à votre cas d'usage.

### Comment préparer vos données ?

#### 1 Bien comprendre ce qui est attendu

C'est la première étape incontournable de votre IA. Pour chaque nouvelle tâche, fonctionnalité, projet... il importe d'échanger avec les métiers pour comprendre précisément ce qu'ils attendent. Dès lors, à vous d'identifier les cas complexes et de les checker avec la MOA pour être certain de bien cadrer les objectifs souhaités.

#### 2 S'assurer de la qualité des données

A-t-on du bruit dans nos données (c'est-à-dire des informations dont on n'a pas besoin pour construire le modèle), des doublons, des informations incomplètes, des erreurs... ? Les big data imposent de filtrer pour avoir les bonnes performances au risque de rencontrer des problèmes de classification. Ce qui est informatif, c'est ce qu'on ne sait pas. En ce sens, la sélection des features, c'est-à-dire des informations utiles, permet de supprimer les données inutiles.

#### 3 Vérifier que l'on dispose de suffisamment de données

L'intelligence artificielle implique un grand volume de données. Dans le cas inverse, il est alors possible d'utiliser certaines techniques pour augmenter sa base de données, comme la Data Augmentation. Il est également possible, dans certains cas, d'affecter des valeurs sur la base d'échantillons similaires, ou simplement de remplacer les données pour les projets moins complexes.

#### 4 Labéliser les données (en apprentissage supervisé)

N'oubliez pas de prendre en compte l'évolution des données dans le temps : vous devez donc travailler avec des données actuelles et passées, et prévoir comment elles peuvent évoluer dans les années à venir.

### Un exemple de préparation des données

Il existe de nombreux exemples de préparation, comme la stratification qui permet de créer des échantillons d'apprentissage équilibrés. En réalité, tout dépend du type de données : texte, image, vidéo, etc. Si on a des images, le risque majeur est d'avoir des matrices qui font des centaines de milliers de lignes, soit des giga-octets de données, ce qui peut compliquer le traitement. Il faut alors redimensionner les données, prétraiter les images de mauvaise qualité et les sauvegarder sur un format qui prend moins d'espace.

## [Question #1]

### » L'avis de l'expert

Vous devez prendre soin de bien choisir à chaque fois les informations les plus pertinentes et de gérer les données manquantes. Si on prend l'exemple de la santé : vous devez identifier les variables qui influent véritablement sur la maladie, comme l'âge ou le statut socio-professionnel, pour ne retenir que les données dont vous avez véritablement besoin. C'est pourquoi il est intéressant de mener les deux étapes de préparation et d'analyse de front : dans la 1re, j'identifie les informations manquantes, et dans la 2e, je sélectionne les plus pertinentes au regard de mon cas d'usage.



**Amady Ba**  
Consultant  
Data Scientist  
& Machine  
Learning  
Engineer

Le manque de données ou les problèmes de qualité des données représentent

# 20%

**des principaux goulots d'étranglement à l'adoption de l'IA.**

Source : [Rapport sur l'adoption de l'IA en entreprise](#), O'Reilly, 2022



## Trouver quel type de modèle tester

Le rôle du data scientist consiste à identifier le meilleur modèle à utiliser pour un cas d'usage donné. Comment ? À cette fin, plusieurs techniques sont possibles.

### Comment choisir son modèle algorithmique ?

#### 1 Choisir le modèle le plus simple

Pour choisir quel algorithme tester, optez toujours par le plus simple en premier (par exemple, les arbres de décision, ou les modèles aléatoires et linéaires pour la régression) avant d'aller vers le plus complexe (la forêt aléatoire, les réseaux de neurones) jusqu'à ce que vous trouviez le bon.

#### 2 Comparer les modèles

Un doute ? N'hésitez pas à mettre en place deux algorithmes ou plus, de les tester, et de comparer les résultats pour choisir le plus performant.

#### 3 Valider les métriques utilisées

Pour déterminer quel modèle est le plus performant, vous devez au préalable vous assurer que toutes les parties prenantes travaillent avec la même métrique de scoring du modèle : soit la précision (accuracy), soit la sensibilité ou le rappel (recall). Plus la précision sera élevée, plus le modèle sera performant.

#### Selon Wikipédia

*La précision (ou valeur prédictive positive) est la proportion des items pertinents parmi l'ensemble des items proposés ; le rappel (ou sensibilité) est la proportion des items pertinents proposés parmi l'ensemble des items pertinents. Ces deux notions correspondent ainsi à une conception et à une mesure de la pertinence.*



Pour la classification, vous pouvez également utiliser les techniques de Spécificité et le F1 Score. Pour la régression, il existe la technique R<sup>2</sup> – coefficient de détermination –, le Root Mean Squared Error (RMSE ou Erreur Quadratique Moyenne - EQM), et le Mean Absolute Error (MAE ou Erreur Absolue Moyenne).

## [Question #2]

### » L'avis de l'expert

Tout d'abord, renseignez-vous sur les cas similaires déjà développés pour voir quel modèle algorithmique a été utilisé. De la même manière, inutile de faire une régression sur une image pour laquelle la classification sera bien plus efficace. Il faut donc réfléchir à ce que l'on veut faire. Le deep learning fonctionnera mieux pour des projets complexes. Mais pour des tâches plus simples, le machine learning suffit amplement. Enfin, n'hésitez pas à utiliser les modèles déjà entraînés par les GAFAM. Vous ne disposerez jamais de la puissance de calcul et du volume de données d'un Google. Ces modèles sont faits pour être réentraînés et adaptés à des tâches spécifiques, à l'image de CamemBERT et de FlauBERT issus du modèle BERT et spécifiques au marché français.



Antony  
Arslanyan  
Ingénieur  
Intelligence  
Artificielle



Près de

# 1 000 milliards

**d'opérations par seconde.  
C'est la vitesse de calcul atteinte par  
la solution de contrôle qualité SIAS  
Deeplearning de Clecim à base de  
réseaux de neurones artificiels.**

Source : [L'Usine Nouvelle, juillet 2022](#)

### Les tips de Charbel Fakhry

Responsable du Centre de Compétences Data Science

- Pour des performances lentes mais plus précises, privilégiez les réseaux de neurones, le random forest ou le SVM non linéaire.
- Pour des modèles rapides mais moins performants, optez plutôt pour les arbres de décision, la régression logistique, ou le SVM linéaire.

**Pour vous aider à choisir, sachez que pour les réseaux de neurones sont majoritairement utilisés pour les données non structurées et les big data.**

## Le prototype

Avant de passer le modèle en production, vous devez tester ses performances dans conditions les plus proches possibles de la réalité. Objectif : détecter les erreurs avant de le déployer à grande échelle.

### Comment tester son modèle efficacement ?

#### 1 Où effectuer son apprentissage : sur un PC en local ou dans le cloud ?

Si vous possédez des machines puissantes, vous pouvez alors tester vos modèles en local. Vous n'aurez pas de frais supplémentaires à payer. En revanche, si votre modèle est complexe et nécessite d'importants volumes de données, vous serez alors plus limités.

Si vous souhaitez en revanche bénéficier d'un environnement de travail plus configurable, plus rapide et plus flexible, privilégiez des machines virtuelles dans le cloud sur des serveurs dédiés. Mais attention au coût qui peut vite grimper selon la configuration souhaitée et à la sécurité des données sensibles.

#### 2 Tester la fonctionnalité du modèle algorithmique

Quand un modèle semble suffisamment robuste, testez-le dans l'application finale mais dans un environnement de préproduction pour corriger les bugs ou les biais.

#### 3 Comprendre les erreurs

En cas d'erreur, il est indispensable de comprendre pourquoi le modèle les a commises. Donc n'hésitez pas à retester et à réentraîner, voire à réinjecter de nouvelles données. Cette étape d'explicabilité du modèle est fondamentale à la fois pour comprendre les décisions prises par le modèle.

### Les erreurs les plus fréquentes

- **Céder à la facilité** : les data scientists sont souvent tentés de choisir le dernier modèle tendance même s'il n'est pas aligné avec leurs besoins. Résultat, le test ne sera pas probant et il faudra recommencer avec un algorithme différent.
- **Mettre en place une architecture avant le cas d'usage** : c'est tout l'inverse qu'il faut faire.
- **Tester le modèle sur des données non représentatives** : vous aurez alors des résultats décorrélés de la phase d'apprentissage du modèle.
- **Négliger la préparation et l'analyse des données** : beaucoup veulent aller trop vite et vont directement à l'étape construction du modèle. Ils sont donc obligés d'itérer à chaque fois, soit une perte de temps énorme.
- **Mal faire son jeu de test** : attention à ce que votre jeu de test soit suffisamment complet et équilibré pour s'appliquer correctement à la vie réelle.
- **Mélanger jeu de testing et jeu d'entraînement** : si vous testez des images que le modèle connaît déjà, le résultat sera nécessairement meilleur (et donc biaisé).

## [Question #3]

### L'avis de l'expert

Dans certaines situations, il est très difficile d'obtenir la vérité terrain : le recueil de données peut s'avérer très complexe car il implique des accords avec de multiples parties prenantes, impose de nombreux déplacements, etc. Résultat, certaines organisations déploient parfois une vérité terrain... sans aller sur le terrain ce qui rend les résultats approximatifs. Autre écueil fréquent : ne pas disposer de la puissance suffisante pour exploiter la masse de données à disposition. Résultat, il est parfois nécessaire de retester sur une nouvelle plateforme plus performante, soit une perte de temps conséquente. Les capacités matérielles représentent un point critique de l'industrialisation de l'IA.



**Abdelkader Hamadi**  
Ingénieur  
Chercheur en  
Data Science

Seules

# 12%

**des entreprises en 2022 utilisent l'IA à un niveau de maturité leur octroyant un avantage concurrentiel significatif.**

Source : ["The Art of AI Maturity: Advancing from Practice to Performance", Accenture, 2022](#)



# Conclusion

**Les entreprises ont pris conscience de l'importance d'industrialiser leurs modèles IA pour acquérir un atout concurrentiel majeur dans un contexte de plus en plus volatile, complexe et incertain. Pour franchir ce cap de la mise à l'échelle et dépasser le stade des PoC, elles doivent néanmoins encore progresser dans leur méthodologie et leur approche de l'intelligence artificielle. Faut-il déployer le modèle sur un serveur local ou plutôt dans le cloud ? Comment mettre en œuvre l'auto-apprentissage pour améliorer notre modèle au fur et à mesure ? Où placer ce modèle dans l'architecture globale ? Peut-on faire confiance aux GAFAM ?**

Autant de questions qui freinent toujours l'adoption de l'IA malgré ses nombreux bénéfices. Heureusement, les data scientists peuvent aujourd'hui compter sur des solutions performantes pour simplifier leurs tâches et accélérer le passage en production, à l'image d'Huggingface qui permet d'entraîner facilement des NLP en quelques lignes de codes, ou de Dall-e qui permet de générer des visuels très facilement et de manière très poussée.

Reste néanmoins le problème de savoir comment adresser la complexité inhérente à l'intelligence artificielle. Or la technologie ne peut suffire à résoudre à elle seule l'ensemble des points de difficulté. Sur la base des démarches DevOps et DataOps, les équipes doivent repenser leur organisation pour tendre vers plus d'agilité, de flexibilité et d'apprentissage continu. C'est toute la promesse portée par le MLOps.

## Meritis, le talent d'aller plus loin.

### CONSEIL, PILOTAGE ET DÉVELOPPEMENT IT

Meritis est une société de conseil en transformation des Systèmes d'Information et Organisations.

#### → Notre approche ?

Accompagner nos clients sur l'ensemble de la chaîne de valeur : cadrage personnalisé, pilotage & développement applicatif pour les projets IT.

#### → Notre mission ?

Connecter les meilleurs talents au service de la transformation numérique pour donner un temps d'avance aux entreprises.

**Nos +800** consultants vous accompagnent avec agilité dans tous vos projets de transformation digitale. Un seul objectif : vous emmener plus loin.

## Nos implantations en France

Nous sommes présents à Paris, Nice Sophia Antipolis, Aix-en-Provence, Nantes et Montpellier.

Nous intervenons également sur Bordeaux, Lille, Rennes, et Toulouse.



## **Nous contacter**

Un projet, une question, vous souhaitez en savoir plus ?

Contactez-nous ! Nos équipes d'experts sont votre disposition pour répondre toutes vos questions.

**NOUS CONTACTER**

**NOUS REJOINDRE**

**in**

**f**



**meritis.fr**

