

# ANALYSER UN BESOIN/ UNE DEMANDE EN IA

MSPR 1

Project by

**Abdelkim BENKIRANE**  
**Mohamed BARHAMI**  
**Hamza Oujja**  
**Mohamed AIT ABDELLAH**

School

**EPSI M1 IA**

Our Topic

**Déploiement d'une solution IA de  
recommandation, segmentation  
et prédiction des ventes pour la  
marque Amazing Basics**

# Sommaire

Introduction	1
--------------	---

Cahier de charges	2
-------------------	---

• Présentation d'Amazing	2
--------------------------	---

• Problématique	2
-----------------	---

• Analyse des données disponibles (Audit initial)	5
---	---

• Objectifs fonctionnels	5
--------------------------	---

• Stratégie Marketing pilotée par IA	6
--------------------------------------	---

• KPI de Suivi du Projet	6
--------------------------	---

• Roadmap et Étapes Clés du Projet	6
------------------------------------	---

• Architecture Data & Processus Métier	7
--	---

• Besoin (Cadrage non-technique)	7
----------------------------------	---

• Audit des Données Disponibles	8
---------------------------------	---

• Étude de Faisabilité : Risques et Contraintes	8
---	---

• Justification du Choix de Projet Data Science	9
---	---

• Évaluation RGPD & Actions Recommandées	9
--	---

<b>Audit des données disponibles</b>	<b>10</b>
<b>Analyse de faisabilité et contraintes</b>	<b>14</b>
<b>Justification de la solution IA retenue</b>	<b>16</b>
<b>Veille technologique, sectorielle, concurrentielle et réglementaire</b>	<b>19</b>
<b>Indicateurs de performance (KPI) et plan de suivi</b>	<b>22</b>
<b>Étude des outils et technologies retenus</b>	<b>24</b>
<b>Roadmap du projet</b>	<b>27</b>
<b>Conformité RGPD et gouvernance des données</b>	<b>28</b>
<b>Évaluation des moyens prévisionnels et estimation budgétaire</b>	<b>30</b>
<b>Note de faisabilité synthétique</b>	<b>31</b>
<b>Conclusion</b>	<b>33</b>

# Introduction

La révolution numérique redéfinit en profondeur les pratiques commerciales, et l'intelligence artificielle (IA) s'impose comme un levier stratégique majeur dans cette transformation. Dans un contexte où les attentes des consommateurs évoluent rapidement, l'IA offre des perspectives inédites pour mieux comprendre les comportements d'achat, personnaliser les offres, et optimiser l'efficacité des actions marketing.

La problématique centrale que nous abordons est donc la suivante : comment exploiter l'intelligence artificielle pour personnaliser l'expérience client et stimuler les ventes, tout en garantissant l'interprétabilité, la conformité réglementaire et l'intégration opérationnelle de la solution ?

Ce cahier des charges s'inscrit dans cette dynamique d'innovation, en se concentrant sur l'usage de l'intelligence artificielle pour améliorer les performances commerciales de la marque Amazing Basics, en particulier sur les catégories en perte de vitesse telles que le divertissement. L'objectif principal est de concevoir une solution data-driven, fondée sur des modèles prédictifs et de recommandation, afin d'augmenter la pertinence des campagnes marketing, renforcer la fidélisation client, et maximiser le retour sur investissement.

L'IA représente une opportunité stratégique pour Amazing : en combinant puissance analytique et adaptation en temps réel, elle permet de dépasser les approches traditionnelles basées sur des segments fixes ou des promotions génériques. Cependant, cette transformation soulève également des enjeux complexes : gouvernance des données, respect du RGPD, lisibilité des modèles, et appropriation par les équipes non techniques.

Ce cahier des charges propose une démarche structurée pour cadrer, concevoir et déployer une solution IA sur mesure. Il est organisé autour de plusieurs dimensions clés : analyse des besoins métiers, audit des données disponibles, choix méthodologiques, planification, contraintes, faisabilité et conformité. L'ambition est d'apporter aux décideurs une vision claire, pragmatique et orientée résultats, pour tirer pleinement parti du potentiel de l'intelligence artificielle dans un cadre maîtrisé et responsable.



# Présentation de l'Entreprise : Amazing

Amazing est une marketplace en ligne internationale qui propose une très large gamme de produits allant de l'électronique à la mode, en passant par les accessoires pour la maison, les produits de beauté, et bien d'autres catégories. L'entreprise est un acteur mondial majeur du e-commerce, présent dans de nombreux pays, avec une plateforme performante, une logistique optimisée, et un modèle axé sur la satisfaction client. Une partie significative de son chiffre d'affaires provient de sa marque propre : Amazing Basics, qui regroupe des produits simples, pratiques et à bas prix.

Cette marque a pour objectif de proposer des articles essentiels à tous les types de consommateurs. Cependant, ces derniers mois, Amazing a constaté une diminution de l'engagement client et une baisse des ventes, notamment dans la catégorie "divertissement". Cette situation s'explique par plusieurs facteurs : l'inflation, une concurrence agressive, une saturation du marché, ainsi qu'une faible personnalisation de l'offre actuelle.

## Problématique

Malgré un positionnement solide et une base de clients massive, Amazing fait face à un recul de 15 % des ventes sur les produits Amazing Basics au cours des deux derniers trimestres. Cette baisse est accentuée sur les produits liés au divertissement et aux accessoires technologiques, autrefois piliers de la croissance.

Les initiatives marketing classiques (newsletters génériques, promotions saisonnières, etc.) ne suffisent plus à stimuler les ventes. Il est aujourd'hui nécessaire d'évoluer vers des actions plus ciblées, personnalisées et intelligentes. L'objectif est de tirer parti des données clients et des technologies d'IA pour restaurer la croissance et améliorer significativement les ventes.

L'enjeu principal est donc : comment exploiter les données internes de manière intelligente afin de relancer les ventes de la marque Amazing Basics, en optimisant le ciblage marketing, l'expérience utilisateur et la prédiction des besoins client ?



# Note de cadrage du projet

## Contexte général

- Amazing est une marketplace internationale reconnue, dont une part significative du chiffre d'affaires repose sur sa marque interne "Amazing Basics". Depuis les derniers épisodes d'inflation, la marque connaît une baisse préoccupante des ventes, notamment sur les produits liés au divertissement. Ce recul s'explique par plusieurs facteurs convergents : une offre peu personnalisée, une saturation du marché, la montée en puissance de concurrents plus agiles, et un changement de comportement des consommateurs.
- Par ailleurs, l'entreprise possède une infrastructure data mature (data lake cloud), une équipe marketing structurée, et une volonté stratégique d'intégrer l'intelligence artificielle comme levier d'innovation commerciale. Dans ce contexte, l'IA apparaît non plus comme une option, mais comme une nécessité compétitive pour anticiper les besoins clients, ajuster les prix, personnaliser les contenus, et automatiser les actions marketing.
- Le présent projet s'inscrit donc dans une dynamique de modernisation des pratiques marketing et de revalorisation de la donnée client à travers une approche data-driven fondée sur des algorithmes prédictifs, interprétables et exploitables opérationnellement.

## Besoin métier reformulé

Les équipes marketing expriment un besoin clair et structurant : relancer les ventes de la marque Amazing Basics en comprenant mieux les profils clients, leurs comportements d'achat, et leurs attentes implicites. Le marketing traditionnel — newsletters génériques, promotions standardisées — a atteint ses limites. L'entreprise souhaite s'appuyer sur l'IA pour :

- Segmenter finement sa base client
- Proposer des recommandations personnalisées en temps réel
- Mieux cibler ses campagnes promotionnelles
- Prédire les ventes pour ajuster les stocks et les prix

Le besoin ne se limite pas à une dimension technique, il est opérationnel et stratégique : améliorer la performance commerciale par une personnalisation intelligente et mesurable.

# Objectifs opérationnels

Les objectifs fixés sont concrets, quantifiables, et alignés sur les priorités de l'entreprise. Ils servent de base à l'évaluation de l'impact de la solution IA déployée :

- Augmenter le taux de conversion des visiteurs en acheteurs d'au moins 10 %
- Élever le panier moyen grâce à la vente croisée et à la pertinence des recommandations (+8 %)
- Optimiser le ROI marketing (+15 %) en réduisant les dépenses inefficaces
- Réduire le taux de rebond sur les pages produit (-20 %)
- Identifier des segments clients à fort potentiel pour des campagnes dédiées
- Automatiser partiellement la création de contenu marketing personnalisé

## Plus-value attendue

La solution IA proposée permettra à Amazing de passer d'un marketing de masse à un marketing de précision, capable d'anticiper les besoins plutôt que d'y répondre tardivement.

Cette plus-value se décline à plusieurs niveaux :

- Business : relance des ventes, fidélisation, meilleure rentabilité des campagnes
- Data : valorisation des données clients, meilleure gouvernance et meilleure qualité
- Organisationnel : montée en compétences des équipes marketing et BI sur l'IA
- Stratégique : différenciation face à la concurrence grâce à une personnalisation en temps réel
- Éthique : IA explicable, respectueuse des données personnelles et conforme au RGPD

En combinant des modèles robustes (XGBoost, LightFM, KMeans) avec une restitution claire (dashboards, API), le projet vise une intégration fluide et une adoption rapide par les équipes métier.



# Analyse des Données Disponibles

L'équipe Data d'Amazing met à disposition une grande quantité de données, structurées dans un Data Lake cloud. Ces données couvrent toutes les dimensions de l'expérience client :

- Catalogue des produits : noms, prix, catégories, stock, descriptions.
- Commandes clients : produits achetés, date, fréquence, valeurs paniers.
- Comportement de navigation : pages visées, temps de session, clics, rebonds.
- Base clients anonymisée : données d'inscription, historique, segments.
- Actions marketing passées : promotions, newsletters, codes réductions.

Ces données sont riches et multi-sources. Elles permettent d'imaginer des modèles complexes comme :

- Des systèmes de recommandation personnalisée (filtrage collaboratif, deep learning)
- Des modèles de scoring clients (segmentation RFM, clustering)
- Des prédictions de ventes selon les périodes, les comportements ou les profils

## Objectifs Détaillés du Projet

Augmentation du chiffre d'affaires de la marque Amazing Basics de 25 % d'ici 12 mois, via des recommandations intelligentes et des actions ciblées

- Amélioration du taux de conversion des visiteurs en acheteurs (+10 %)
- Hausse du panier moyen (+8 %) grâce à la vente croisée à travers des suggestions IA
- Optimisation du retour sur investissement marketing (ROI campagnes +15 %)
- Réduction du taux de rebond de 20 % sur les pages produits

Pour cela, la solution doit :

- **Identifier les profils d'utilisateurs types et leurs besoins**
- **Détecter les produits sous-performants et ceux à fort potentiel**
- **Personnaliser les recommandations produits et les e-mails**
- **Prévoir la demande pour adapter les prix et les stocks**



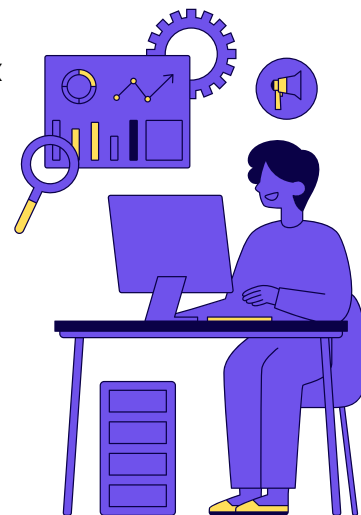


# Stratégie Marketing Pilotée par l'IA

La relance des ventes passera par une réorientation du marketing vers la personnalisation :

- Segmentation IA : découper les clients en groupes comportementaux (jeunes technophiles, parents actifs, clients inactifs, etc.)
- Campagnes ciblées : envoi de newsletters et offres personnalisées à fort taux de réponse
- Recommandations temps réel : sur le site ou par email, selon l'historique utilisateur
- Dynamic pricing : moduler les prix selon le moment ou le profil client
- Cross-selling & up-selling automatisé : suggestion de produits associés

Cette stratégie sera testée avec des groupes de clients et optimisée selon les résultats (A/B testing).



## —● KPI de Suivi du Projet

- ↑ Taux de conversion (avant/après IA)
- ↑ Panier moyen
- ↑ Taux de clics sur recommandations personnalisées
- ↑ Taux d'ouverture des campagnes emails ciblées
- ↑ Évolution des ventes par catégorie Amazing Basics



## Roadmap et Étapes Clés du Projet ●—

La feuille de route du projet a été pensée de manière progressive et itérative, pour garantir à la fois rigueur méthodologique et flexibilité d'adaptation. Elle se structure autour de cinq grandes phases allant de l'analyse initiale jusqu'au déploiement opérationnel et à l'amélioration continue. Chaque étape est définie avec des objectifs clairs, des livrables concrets, et des délais réalistes tenant compte des ressources disponibles, tant du côté de l'équipe data que de celle du commanditaire. Cette roadmap permet d'assurer un suivi fluide du projet, de prioriser les actions critiques, et d'assurer une montée en valeur incrémentale au fil du temps..

# Architecture Data & Processus Métier

Étapes clés à représenter dans un schéma métier :

- Collecte des données clients, navigation et produits
- Traitement, nettoyage et agrégation dans le Data Lake
- Modélisation (modèles IA pour recommandation, scoring, prédiction)
- Déclenchement des actions marketing (emails, recommandations web)
- Suivi des KPIs en temps réel via dashboard BI



## —● Conformité RGPD et Éthique

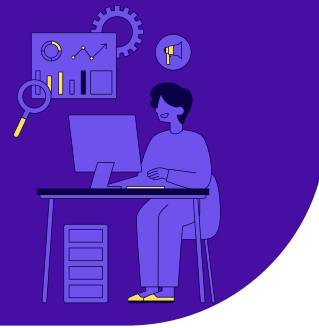
- Toutes les données utilisées doivent être anonymisées et stockées selon les normes européennes
- Le client doit pouvoir accéder, rectifier ou supprimer ses données sur demande
- L'algorithme ne doit pas prendre de décisions discriminatoires (ex. par sexe ou origine)
- Ajout de mécanismes de contrôle humains (interprétabilité des modèles)
- Formation de l'équipe marketing sur les implications légales



## —● Outils et Technologies Recommandées

- Data Engineering : AWS S3, AWS Glue, Snowflake, Airflow
- Modélisation IA : Python, Scikit-learn, XGBoost, TensorFlow, LightFM
- Visualisation : Power BI, Streamlit pour dashboard interactif
- Déploiement : Docker, FastAPI, CI/CD GitHub Actions
- Surveillance : Prometheus, Grafana pour monitoring

# Besoin et objectifs opérationnels



Ce projet vise à répondre à un besoin clairement exprimé par les équipes d'Amazing : relancer les ventes de la marque Amazing Basics grâce à une meilleure connaissance client et à une personnalisation accrue de l'expérience d'achat. L'objectif n'est pas uniquement technique, mais avant tout opérationnel : proposer une solution intelligente, simple à exploiter par les équipes marketing, et intégrée aux outils existants. Il s'agit d'allier innovation en intelligence artificielle et accessibilité métier.



## —● Audit des données disponibles

Avant toute phase de modélisation, une évaluation rigoureuse de la qualité, de la pertinence et de la complétude des données disponibles est nécessaire. Cet audit permet d'identifier les forces et les limites des sources actuelles (navigation, ventes, campagnes marketing), et de définir les actions préalables indispensables pour garantir la fiabilité des analyses et des modèles à venir.

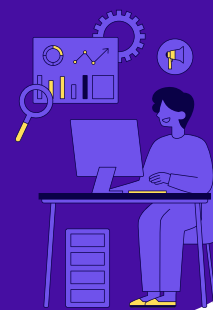
## —● Analyse de faisabilité, contraintes et risques

La mise en œuvre du projet doit tenir compte de plusieurs contraintes techniques, humaines et réglementaires. Une analyse de faisabilité permet d'anticiper les principaux risques liés au calendrier, à la complexité des modèles ou encore à la conformité RGPD. Cette évaluation conditionne la réussite du projet, en assurant un équilibre entre ambition technologique et réalisme opérationnel.

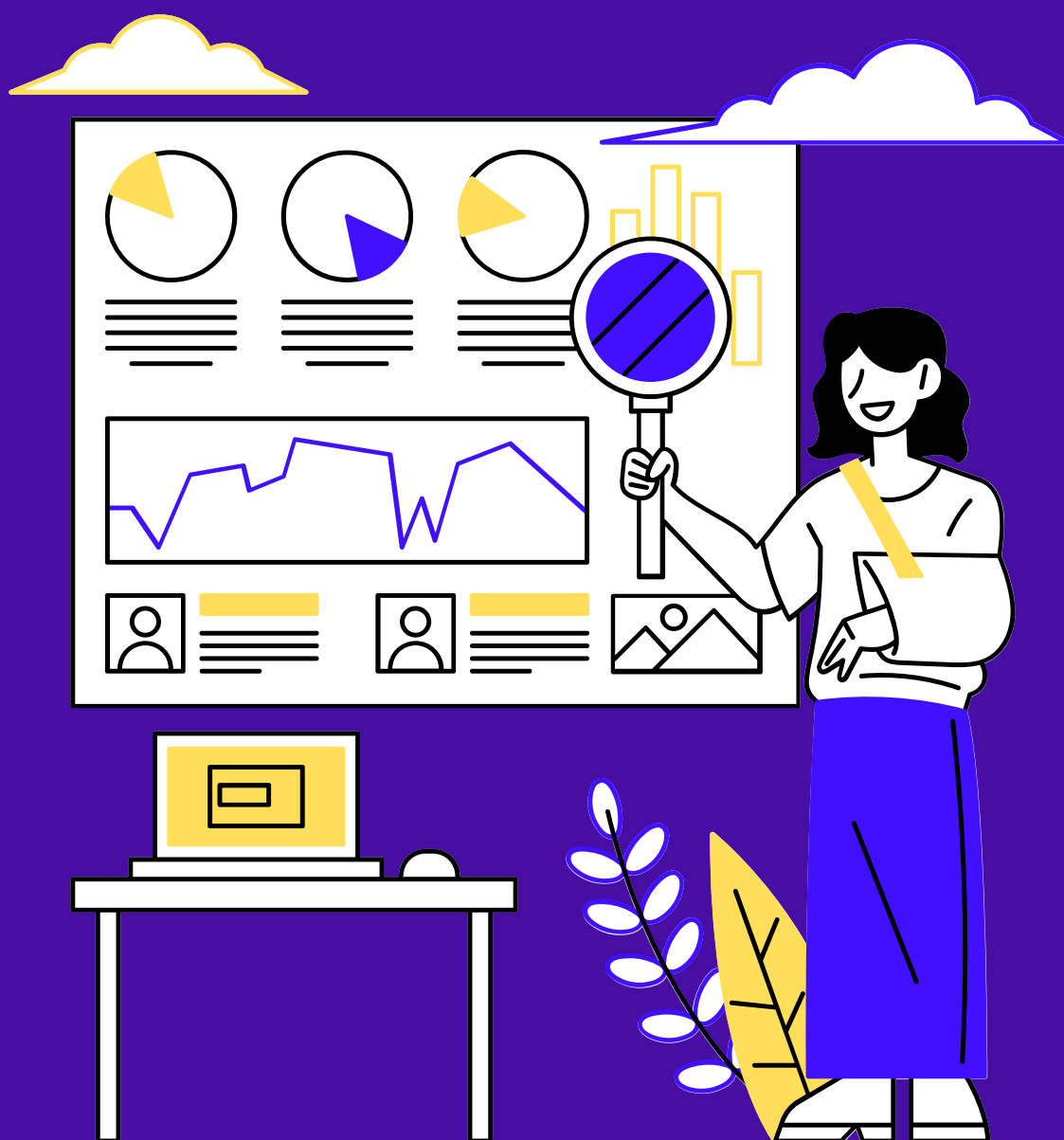
## —● Pertinence de la solution proposée

Le recours à une approche data science se justifie pleinement au regard des enjeux identifiés. Cette démarche permet de répondre de manière ciblée et personnalisée aux attentes des clients tout en optimisant les actions marketing. Le projet proposé s'inscrit dans une logique de création de valeur mesurable, durable et adaptée aux nouvelles dynamiques du marché.

# Conformité RGPD et gouvernance des données



L'ensemble du projet sera conduit dans le strict respect du Règlement Général sur la Protection des Données (RGPD). Au-delà des obligations légales, la gouvernance des données constitue un levier de confiance auprès des utilisateurs. Des actions concrètes seront mises en place pour garantir la transparence, la sécurité et l'équité des traitements de données tout au long du projet.



# Audit de la qualité des données

## Objectif de l'audit

Avant toute démarche de modélisation IA, il est indispensable de réaliser un audit rigoureux de la qualité des données disponibles. Cet audit permet de :

- Évaluer la complétude, la cohérence et la fiabilité des données.
- Identifier les lacunes structurelles (données manquantes, colonnes bruitées, duplications, etc.).
- Évaluer la pertinence des données au regard des objectifs IA fixés (recommandation, segmentation, prédiction).
- Déterminer les actions de prétraitement nécessaires à la mise en œuvre d'une solution IA fiable et conforme.

## Vue d'ensemble des données

Le jeu de données mis à disposition a été extrait de la plateforme Rees46 et couvre plusieurs mois d'activité e-commerce (octobre 2019 à avril 2020). Il est structuré en format .csv.gz et représente plus de 15 Go de logs utilisateur.

Chaque ligne représente un événement utilisateur (consultation, ajout au panier, achat) horodaté, associé à un produit et un identifiant utilisateur pseudonymisé.

## Variables disponibles :

- event\_time : date et heure de l'événement
- event\_type : type d'action (view, cart, purchase)
- user\_id : identifiant pseudonymisé de l'utilisateur
- user\_session : session temporaire
- product\_id, brand, category\_code, price

## • PROBLÈMES DÉTECTÉS



### TYPE DE PROBLÈME

### DESCRIPTION DÉTAILLÉE

#### Valeurs manquantes

Absence fréquente de category\_code, brand ou price. Certains prix sont à 0 ou null.

#### Valeurs aberrantes

Quelques prix excessifs (>10 000€) non représentatifs du marché cible.

#### Colonnes inutilisables

Aucune donnée socio-démographique (âge, sexe, localisation).

#### Données absentes

Aucune trace des actions marketing (emails envoyés, promotions reçues).

#### Sessions dupliquées

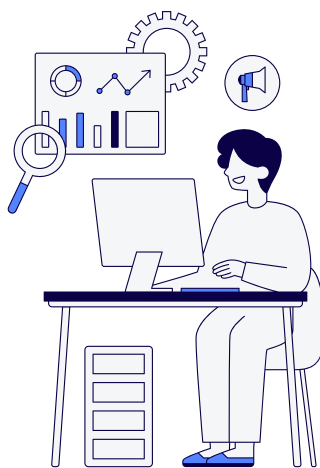
Même user\_session associé à des event\_time incohérents, nécessitant un filtrage.

## • DONNÉES EXPLOITABLES PAR OBJECTIF IA

OBJECTIF IA	VARIABLES PERTINENTES DISPONIBLES
Recommandation produits	user_id, product_id, event_type, user_session, price
Segmentation comportementale	Fréquence des sessions par user_id, taux de conversion, récurrence
Prédiction des ventes	event_time, event_type=purchase, price, category_code

## • ACTIONS CORRECTIVES RECOMMANDÉES

ACTION DE NETTOYAGE	DESCRIPTION
Traitement des prix nuls ou aberrants	Supprimer ou imputer selon la médiane des produits similaires
Normalisation des catégories	Compléter les category_code manquants via regroupement par product_id
Filtrage des sessions invalides	Exclure les user_session de moins de 2 événements ou incohérentes
Agrégation des événements	Regrouper les interactions par utilisateur-session pour créer un profil
Création de variables dérivées	RFM, fréquence d'achat, récence, valeur totale panier, etc.



## • CONFORMITÉ RGPD ET QUALITÉ ÉTHIQUE

Les données sont pseudonymisées (user\_id, user\_session) et ne contiennent pas d'éléments directement identifiants. Toutefois, la qualité de certaines colonnes peut biaiser les modèles (par exemple, sur-représentation d'une catégorie de produits ou de comportements).

Il est donc recommandé :

- **D'effectuer un audit des biais algorithmiques sur les segments de clientèle.**
- **De veiller à une juste représentativité des comportements dans l'entraînement des modèles.**
- **D'intégrer un mécanisme d'explication des décisions automatisées dès la phase de prototypage.**

## Conclusion de l'audit

Les données mises à disposition sont riches et structurées, ce qui constitue une base solide pour développer une solution IA de recommandation, segmentation et prédiction. Néanmoins, plusieurs actions de préparation sont nécessaires pour garantir :

- La fiabilité des résultats algorithmiques,
- La conformité aux bonnes pratiques en data science,
- Et la cohérence métier avec les objectifs commerciaux d'Amazing.
- 

Une fois nettoyées et enrichies, ces données permettront d'alimenter des modèles IA performants et d'atteindre les objectifs fixés par l'entreprise dans des délais réalistes.



# Analyse des risques

## Enjeux de l'analyse des risques

Dans un projet d'intelligence artificielle à visée commerciale, comme celui porté par Amazing, il est fondamental d'anticiper les risques pour sécuriser à la fois la qualité des résultats, leur adoption par les métiers, et leur conformité réglementaire. La méthode EY (Ernst & Young) propose une grille structurée d'analyse croisant :

- Les objectifs IA visés
- Les risques spécifiques liés à chaque usage
- Et les actions préventives ou correctives associées

L'objectif est d'assurer la maîtrise du périmètre, la robustesse des algorithmes, et la sérénité des parties prenantes tout au long du cycle de vie du projet IA.

## • CONTRAINTES TECHNIQUES

**Infrastructure nécessaire** : le traitement de données volumineuses (>15 Go) et l'entraînement de modèles IA nécessitent une infrastructure scalable. Amazing dispose déjà d'un Data Lake cloud, ce qui est un atout.

**Prétraitement important** : les données brutes (fichiers .csv.gz) exigent des étapes de nettoyage, d'agrégation, et de gestion des valeurs manquantes avant toute modélisation.

**Absence de données contextuelles** (stock, marketing, données client enrichies) : la prédiction reste limitée aux comportements utilisateurs enregistrés.

**Outils nécessaires** : Python, scikit-learn, LightFM, XGBoost, ou encore FastAPI pour le déploiement.

## • CONTRAINTES HUMAINES

**Adoption par les équipes métiers** : les équipes marketing devront être formées à l'exploitation des résultats IA (tableaux de bord, modèles de recommandation, scores de segmentation).

**Compréhension des modèles** : certains algorithmes (Random Forest, XGBoost) peuvent être vus comme des "boîtes noires". Il sera nécessaire de fournir des visualisations ou explications simplifiées pour faciliter leur appropriation.

**Résistance au changement** : il faut anticiper une phase d'accompagnement pour que les équipes passent d'un marketing générique à un marketing basé sur la donnée.





## • CONTRAINTES RÉGLEMENTAIRES (RGPD)

Les données utilisées sont pseudonymisées (`user_id`, `user_session`) et ne contiennent pas d'informations personnelles directes, ce qui limite les risques.

Cependant, l'utilisation de ces données à des fins de profilage nécessite :

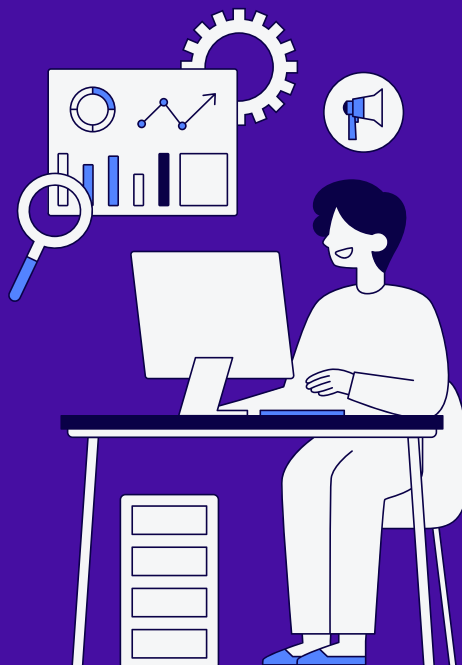
- Une transparence sur les traitements
- Une évaluation des biais possibles
- L'intégration de mécanismes de contrôle (droits d'accès, rectification, suppression)

L'analyse de la conformité devra être validée avec un DPO (Data Protection Officer) en cas de mise en production réelle.

## • LIEN AVEC LA GOUVERNANCE PROJET

Les risques techniques et humains font l'objet d'un suivi actif par le chef de projet IA, en lien avec :

- Le Data Engineer pour la qualité des flux de données,
- Le Data Scientist pour l'explicabilité et la performance des modèles,
- Le Référent marketing pour l'adoption par les utilisateurs métiers,
- Le DPO ou juriste RGPD pour garantir la conformité légale.



# Analyse des risques

## • CAPACITÉ DU DATA LAKE



Le Data Lake mis à disposition par Amazon constitue un socle robuste pour le stockage massif et historisé des données clients, produits et interactions. Toutefois, il ne suffit pas à lui seul pour exécuter des traitements temps réel. Pour répondre aux exigences de recommandation dynamique ou d'interactions personnalisées à la volée (ex. : suggestions produit pendant la navigation), une infrastructure de traitement distribué en streaming (type Spark Streaming, Flink, ou AWS Kinesis) sera nécessaire. Ce décalage entre stockage à froid (Data Lake) et traitement à chaud (inférence temps réel) constitue un risque d'inadéquation technique, à anticiper dès la conception.

## • COMPATIBILITÉ AVEC LA RECOMMANDATION EN TEMPS RÉEL



La logique de recommandation en temps réel suppose :

- une capacité à ingérer et traiter des événements utilisateurs en continu,
- un accès optimisé aux modèles IA déjà entraînés,
- une latence inférieure à quelques centaines de millisecondes.

Or, dans l'architecture actuelle, aucune couche de traitement temps réel n'est explicitement intégrée, ce qui pourrait limiter les cas d'usage live (produits suggérés à la volée, mise à jour dynamique du panier, etc.). Il est donc recommandé de prévoir une brique complémentaire de type FastAPI couplée à un cache (Redis, Faiss, etc.), pour le scoring à la demande.

## • SUIVI UTILISATEUR VIA USER\_ID ET USER\_SESSION : ENTRE OPPORTUNITÉ ET VIGILANCE

Les identifiants `user_id` (utilisateur pseudonymisé) et `user_session` (session temporaire) permettent de reconstituer les parcours utilisateurs, qu'ils soient enregistrés ou non, ce qui constitue un atout majeur pour l'analyse comportementale. Cependant, cela implique une grande prudence vis-à-vis du RGPD, notamment en matière de profilage (article 22). Il conviendra de distinguer clairement les usages anonymes et connectés, et d'éviter toute confusion entre identifiants techniques et données personnelles.

## • NE PAS CONFONDRE LES NIVEAUX D'ARCHITECTURE : STOCKAGE, TRAITEMENT, EXPLOITATION

Il est essentiel de ne pas placer au même niveau les briques techniques du projet. Le stockage (ex. : AWS S3, Data Lake) assure uniquement la centralisation et l'historisation des données. Le traitement repose sur des outils comme Airflow, XGBoost ou LightFM, qui permettent la préparation, l'entraînement et le scoring en mode batch. Enfin, l'exploitation concerne les interfaces et services (FastAPI, Streamlit, Redis) qui permettent la restitution des résultats, l'interaction utilisateur, et la recommandation en temps réel. Chaque couche a un rôle distinct et doit être pensée séparément pour garantir la cohérence et la performance de l'ensemble.

## • ÉVALUATION DES RISQUES (MÉTHODE EY)

OBJECTIF IA	RISQUE POTENTIEL	ACTION PRÉVENTIVE OU CORRECTRICE
<b>Recommandation produits</b>	<ul style="list-style-type: none"> <li>Biais implicites sur certaines catégories (ex. prix, marques)</li> <li>Cold-start utilisateur ou produit</li> <li>Recommandations non pertinentes → désengagement utilisateur</li> </ul>	<ul style="list-style-type: none"> <li>Vérification des biais par segment, filtrage équitable</li> <li>Utilisation de modèle hybride (LightFM), enrichissement par contenu</li> <li>Intégration d'un feedback loop utilisateur, tests A/B</li> </ul>
<b>Segmentation client (RFM)</b>	<ul style="list-style-type: none"> <li>Mauvaise qualité des données → groupes incohérents</li> <li>Segments peu compréhensibles par les métiers</li> <li>Perception de "boîte noire"</li> </ul>	<ul style="list-style-type: none"> <li>Nettoyage rigoureux, exclusion des sessions incomplètes</li> <li>Visualisation claire (cluster map), nommage marketing adapté</li> <li>Restitution pédagogique des clusters, documentation claire</li> </ul>
<b>Prédiction des ventes</b>	<ul style="list-style-type: none"> <li>Saisonnalité non prise en compte</li> <li>Surapprentissage sur certains produits</li> <li>Fluctuation du stock non intégrée dans le modèle</li> </ul>	<ul style="list-style-type: none"> <li>Modèles temporels intégrés (Prophet, XGBoost avec time vars)</li> <li>Validation croisée, régularisation, dropout</li> <li>Enrichissement futur avec données stock</li> </ul>
<b>Adoption par les métiers</b>	<ul style="list-style-type: none"> <li>Manque de confiance dans l'IA ou rejet</li> <li>Difficulté à interpréter les décisions du modèle</li> <li>Résistance à l'intégration dans les pratiques</li> </ul>	<ul style="list-style-type: none"> <li>Co-construction des dashboards, ateliers de restitution</li> <li>Utilisation d'outils XAI (SHAP, feature importance)</li> <li>Sessions de formation marketing &amp; BI, accompagnement terrain</li> </ul>
<b>Conformité RGPD</b>	<ul style="list-style-type: none"> <li>Traitement non justifié ou mal documenté</li> <li>Données ré-identifiables</li> <li>Décisions IA perçues comme discriminantes</li> </ul>	<ul style="list-style-type: none"> <li>Registre de traitement, documentation claire, PIA</li> <li>Pseudonymisation stricte, audit régulier</li> <li>Explicabilité systématique, supervision humaine intégrée</li> </ul>

— Cette grille de lecture a permis d'anticiper les principaux points de vigilance du projet. Elle assure que les choix techniques sont alignés avec les enjeux business et réglementaires, tout en renforçant l'acceptabilité métier.

Chaque risque identifié est associé à une mesure de mitigation claire, qui sera intégrée au plan projet (roadmap et gouvernance) et monitorée pendant les phases de test, restitution et mise en production éventuelle.

# Justification de la solution IA retenue

Afin de répondre efficacement aux besoins exprimés par l'entreprise Amazing, nous avons identifié trois axes fonctionnels prioritaires : la personnalisation de l'expérience utilisateur, la segmentation comportementale de la clientèle, et la prévision de la demande. Pour chacun de ces axes, nous avons sélectionné des modèles d'intelligence artificielle adaptés, en tenant compte des contraintes techniques du projet, de la structure des données disponibles, de la robustesse des algorithmes et de leur degré d'interprétabilité. Cette section présente les modèles choisis, les raisons de leur sélection, ainsi qu'une justification des alternatives écartées.

## • RECOMMANDATION PERSONNALISÉE

La baisse d'engagement constatée sur la marque Amazing Basics appelle une stratégie de personnalisation renforcée. Un système de recommandation permet de proposer aux utilisateurs des produits correspondant à leurs préférences et comportements passés, favorisant ainsi la conversion et la fidélisation.

### Modèle retenu : LightFM

Nous avons choisi l'algorithme LightFM, un système de recommandation hybride combinant :

- le filtrage collaboratif (basé sur les interactions entre utilisateurs et produits),
- et le filtrage par contenu (caractéristiques produits : catégorie, prix, marque).

LightFM est particulièrement adapté à notre dataset, qui contient des événements de type view, cart, purchase, associés à des identifiants produits, utilisateurs, catégories et prix.

Avantages principaux :

- Gère efficacement le problème du cold-start (nouveaux produits ou utilisateurs peu actifs).
- Permet d'exploiter les attributs produits en plus des historiques d'interactions.
- Offre des performances satisfaisantes avec un temps d'entraînement raisonnable.

### Pourquoi pas d'autres modèles ?

- Les approches par réseaux de neurones (type Autoencoders ou GRU4Rec) sont plus complexes, nécessitent davantage de données nettoyées et de puissance de calcul.
- Le filtrage collaboratif pur (ex. SVD) n'utilise pas les attributs produits, limitant la précision dans notre contexte où certaines catégories sont mal renseignées.



## • SEGMENTATION COMPORTEMENTALE

La compréhension des profils clients est essentielle pour adapter les actions marketing. En l'absence de données socio-démographiques, une segmentation basée sur le comportement d'achat s'impose.

### Modèle retenu : K-Means sur une base RFM

Nous avons retenu une approche classique de segmentation RFM (Récence, Fréquence, Montant), sur laquelle nous appliquons le clustering K-Means. Cette méthode permet de regrouper les clients selon leur activité d'achat et de détecter :

- des profils fidèles et actifs,
- des clients à réactiver,
- ou encore des profils occasionnels.

Avantages principaux :

- Algorithme rapide, facilement interprétable et applicable à grande échelle.
- Représentation visuelle claire des clusters.
- Actionnable directement par les équipes marketing.

Pourquoi pas d'autres modèles ?

- DBSCAN est plus adapté à la détection d'anomalies qu'à la segmentation marketing classique.
- Les modèles de mélange gaussien (GMM) sont plus lents et sensibles au bruit, et peu lisibles pour un usage métier.
- Le clustering hiérarchique est difficile à appliquer sur des données volumineuses comme les nôtres.

## • PRÉVISION DES VENTES

Anticiper la demande est un enjeu stratégique pour optimiser les stocks, planifier les campagnes et ajuster les prix. La nature temporelle de nos données justifie une approche de prévision multivariée des ventes.

Modèles retenus :

- **XGBoost (modèle principal)**
- **Random Forest (modèle de benchmark)**

### a) XGBoost

XGBoost est un algorithme de boosting d'arbres de décision, reconnu pour son efficacité sur des données tabulaires complexes.

Avantages principaux :

- Gère bien les relations non linéaires entre les variables (ex. : influence du prix selon la catégorie).
- Offre une grande robustesse même avec des données bruitées ou incomplètes.
- Permet d'identifier les variables les plus influentes dans les prédictions (interprétabilité métier).

Pourquoi pas la régression linéaire ?

La relation entre les variables explicatives et le volume des ventes n'est pas linéaire, ce qui limite la pertinence de modèles trop simples.

c) Random Forest

Enfin, nous avons ajouté le modèle Random Forest comme alternative complémentaire à XGBoost. Il constitue un benchmark robuste pour la régression, avec des résultats souvent comparables à XGBoost en termes de précision, mais plus simple à mettre en œuvre.

Avantages principaux :

- Très bon compromis performance / interprétabilité.
- Facilité d'implémentation sans réglage complexe.
- Moins sensible aux hyperparamètres que les modèles de boosting.

Pourquoi l'inclure malgré XGBoost ?

L'objectif est de comparer plusieurs modèles sur les mêmes indicateurs de performance (RMSE, MAE, etc.) afin de valider le modèle le plus adapté en phase de test.

# Données utiles selon les objectifs IA

OBJECTIF IA	TYPE DE MODÈLE IA	MODÈLES LES PLUS COURANTS	BRÈVE EXPLICATION
Recommandation produits	Filtrage collaboratif ou contenu	-Matrix Factorization (ALS, SVD) -LightFM -Nearest Neighbors (KNN) -Réseaux de neurones (Autoencoders, Deep Learning)	On prédit les produits qu'un utilisateur pourrait aimer en se basant sur ses interactions passées (clics, achats) ou sur les similarités avec d'autres utilisateurs.
Segmentation comportementale	Clustering non supervisé	-K-Means -DBSCAN -Hierarchical Clustering	On regroupe les clients selon leur comportement (fréquence d'achat, récurrence, panier moyen...) pour adapter les actions marketing.
Prédiction des ventes	Régression supervisée / séries temporelles	-XGBoost / LightGBM / CatBoost -ARIMA / SARIMA -Facebook Prophet - RNN / LSTM (Deep Learning)	On prédit le volume ou le montant des ventes futures selon les dates, catégories, prix, etc.

La combinaison des modèles retenus — LightFM, K-Means (RFM), XGBoost, Prophet, Random Forest — permet de répondre de manière complète et cohérente aux objectifs du projet. Elle concilie pertinence métier, performance technique, robustesse face aux données réelles, et capacité à être comprise et exploitée par les équipes opérationnelles d'Amazing.

Ce choix réfléchi garantit une solution IA actionnable, évolutive, et intégrée dans une démarche responsable, en cohérence avec les principes d'éthique, d'explicabilité et de conformité posés dès le début du projet.

# Veille technologique, sectorielle, concurrentielle et réglementaire

La mise en œuvre d'un projet d'intelligence artificielle ne peut se faire sans une compréhension fine de son environnement externe. Dans un secteur aussi évolutif que le e-commerce, il est essentiel de s'inscrire dans une dynamique d'innovation continue, en intégrant les évolutions technologiques, les usages du marché, les pratiques des concurrents et les contraintes juridiques. Cette veille multidimensionnelle alimente les décisions stratégiques du projet et garantit son alignement avec les standards du secteur.

## • VEILLE TECHNOLOGIQUE

Ces dernières années, plusieurs tendances technologiques transforment les usages de l'intelligence artificielle dans les entreprises :

**L'IA générative** : représentée par des modèles comme GPT-4 ou DALL-E, elle permet de générer automatiquement du contenu (textes, images, recommandations, réponses client, etc.). Dans le e-commerce, elle est utilisée pour créer des fiches produits dynamiques, générer des réponses automatisées aux clients ou produire des recommandations ultra-personnalisées.

**Les modèles explicables (XAI – Explainable AI)** : avec des outils comme SHAP, LIME ou ELI5, les entreprises peuvent visualiser les contributions des variables dans les décisions des modèles. C'est un enjeu central dans un contexte RGPD où la transparence est primordiale.

**Le Edge Computing et le MLOps** : les workflows MLOps (Machine Learning Operations) deviennent incontournables pour industrialiser les modèles IA. Ils permettent d'automatiser les phases de versioning, tests, déploiement et surveillance des modèles.

## • Positionnement du projet :

Notre solution IA s'intègre dans cette dynamique technologique, notamment via :

- l'utilisation d'un système de recommandation hybride interprétable (LightFM),
- la sélection de modèles explicables (XGBoost avec importance des variables, Random Forest avec SHAP),
- l'ouverture possible à des outils de monitoring (Streamlit, Prometheus, etc.) pour accompagner le déploiement.



## • VEILLE CONCURRENTIELLE

**Pratiques des leaders du secteur:** Les grandes plateformes e-commerce exploitent l'IA de manière avancée pour personnaliser l'expérience utilisateur et optimiser la rentabilité.

**Amazon :** pionnier du filtrage collaboratif à grande échelle, Amazon utilise un système de recommandation combiné à du deep learning. L'algorithme est optimisé en continu à partir de milliards d'interactions. Amazon utilise également le dynamic pricing (variation de prix en temps réel), basé sur la demande, le profil utilisateur et la concurrence.

**Netflix :** bien qu'opérant dans le secteur du streaming, Netflix est reconnu pour sa recommandation basée sur l'historique utilisateur et ses algorithmes de clustering comportemental, qu'il ajuste pour chaque profil.

**Zalando :** l'entreprise allemande mise sur la personnalisation du catalogue, en croisant comportement utilisateur et tendances de mode. Elle applique aussi le marketing prédictif pour anticiper les produits à promouvoir auprès de chaque segment.

**Spotify et YouTube :** bien qu'orientés média, ces géants exploitent des recommandations contextuelles dynamiques, inspirantes pour la structuration d'un moteur de recommandation e-commerce.

### • Apport au projet :

Ces benchmarks nous confirment que notre choix d'un moteur de recommandation hybride (LightFM) est cohérent avec les pratiques de référence, tout en restant plus léger à déployer.

## • VEILLE SECTORIELLE

**Tendances IA dans le e-commerce:** Le secteur du commerce en ligne évolue rapidement vers une exploitation intensive des données. Voici les principales tendances qui structurent les stratégies IA des entreprises du secteur :

**Personnalisation avancée :** elle devient un standard, allant au-delà de la simple recommandation de produit. Il s'agit d'adapter l'ensemble du parcours client (page d'accueil, suggestions, contenu, prix) en fonction du comportement.

**Marketing prédictif :** les marques utilisent des algorithmes de prédiction pour anticiper les abandons de panier, les comportements de désabonnement, ou encore les besoins latents des utilisateurs. Cela permet de cibler plus intelligemment les campagnes.

**Dynamic Pricing :** les prix ne sont plus fixes, mais évolutifs, en fonction de la demande, du profil utilisateur, de la saisonnalité ou de l'état du stock. Cette pratique devient de plus en plus courante, notamment dans la mode, la tech ou les voyages.

**Segmentation comportementale :** grâce à l'IA, les entreprises délaissent les segments sociodémographiques classiques au profit de clusters dynamiques basés sur les interactions, la récurrence ou la fréquence d'achat.



- **Lien avec notre projet :**

Notre solution IA s'aligne parfaitement avec ces tendances sectorielles : segmentation par K-Means, marketing personnalisé via des clusters RFM, et recommandations ciblées en fonction du comportement utilisateur.

- **VEILLE RÉGLEMENTAIRE**

**Cadre légal et bonnes pratiques: Le déploiement d'un projet IA, même dans un cadre pédagogique, doit être pensé dans le respect des règles de protection des données personnelles et des principes d'éthique algorithmique.**

**RGPD (Règlement Général sur la Protection des Données) :**

- Il impose la minimisation des données utilisées, la finalité explicite des traitements, et le droit d'accès, de rectification ou de suppression pour les utilisateurs.
- Les données utilisées dans notre projet sont pseudonymisées (user\_id, user\_session), ce qui limite les risques de ré-identification.

**IA Act (proposition européenne) :**

- Ce règlement à venir prévoit des niveaux de risques selon les usages de l'IA. Notre projet serait classé en "IA à faible risque", mais il impose tout de même des obligations de transparence, de documentation et de contrôle humain.

**Recommandations de la CNIL :**

- La CNIL encourage l'utilisation de modèles explicables, l'audit des biais algorithmiques, et la mise en place de procédures d'analyse d'impact (PIA) pour certains traitements.

**Mesures prises dans notre projet :**

- Limitation des données à des logs pseudonymisés
- Visualisation des contributions des variables (feature importance, SHAP)
- Sensibilisation à l'explicabilité dans le choix des modèles
- Intégration possible d'une supervision humaine dans le processus de décision

- **CONCLUSION DE LA VEILLE**



L'environnement technologique, concurrentiel, sectoriel et réglementaire dans lequel s'inscrit notre projet confirme la pertinence de la démarche engagée. En mobilisant des modèles IA performants, explicables et respectueux du cadre juridique, notre solution s'inscrit pleinement dans les standards attendus aujourd'hui par les entreprises et les autorités. Cette veille continue permet également d'envisager des pistes d'amélioration futures, telles que l'intégration d'IA générative pour le contenu, ou la mise en œuvre de recommandations en temps réel via des architectures edge/cloud hybrides.

# Indicateurs de performance (KPI) et plan de suivi

Le déploiement d'une solution d'intelligence artificielle n'a de valeur que s'il est mesurable dans sa capacité à produire des résultats concrets. Pour assurer le pilotage du projet et en démontrer l'impact, un ensemble d'indicateurs clés de performance (Key Performance Indicators – KPI) a été défini. Ces indicateurs couvrent à la fois les objectifs marketing, commerciaux et opérationnels liés à l'implémentation de la solution IA.

## • OBJECTIFS DE SUIVI



L'objectif principal de cette phase est de permettre un suivi régulier, structuré et quantifiable de l'efficacité des modèles IA mis en place. Il s'agit d'observer l'évolution des résultats sur une période donnée et de comparer les performances avant et après l'intégration de l'intelligence artificielle dans les processus métiers.

Les indicateurs sélectionnés ont été choisis pour leur pertinence métier, leur mesurabilité dans les données disponibles et leur capacité à refléter directement l'impact de la personnalisation et de la prédiction.

## • PLAN DE SUIVI ET GOUVERNANCE

Pour assurer un pilotage efficace, un plan de suivi structuré sera mis en œuvre dès la phase de déploiement :

**Tableau de bord dynamique :** un dashboard interactif (via Power BI ou Streamlit) permettra de suivre les KPI en temps réel. Il pourra être consulté par les équipes marketing, data et direction.

### Points de contrôle réguliers :

- Un reporting hebdomadaire sur les KPI clés (taux de conversion, panier moyen, rebond)
- Un bilan mensuel incluant des comparatifs avant/après IA
- Une analyse trimestrielle des segments de clientèle et des performances des modèles

### Mise à jour des modèles :

- Les modèles seront réentraînés périodiquement à partir de nouvelles données (approche continue)
- Des mécanismes de surveillance automatique (monitoring des dérives de performance) seront envisagés en phase d'industrialisation

### Boucle de feedback métier :

- Les retours des équipes marketing seront intégrés pour ajuster les actions proposées par les modèles (recommandations, ciblages)
- Un processus d'amélioration continue sera mis en place pour aligner les résultats avec les objectifs business

# KPI retenus

KPI	DÉFINITION	OBJECTIF VISÉ	FRÉQUENCE DE SUIVI
<b>Taux de conversion</b>	Pourcentage de visiteurs ayant effectué un achat	+10 % grâce à une meilleure personnalisation	Hebdomadaire
<b>Panier moyen</b>	Valeur moyenne d'un panier client	+8 % par effet de vente croisée (recommandation IA)	Hebdomadaire
<b>Taux de rebond</b>	Proportion d'utilisateurs quittant le site sans interaction	Réduction de 20 % par des suggestions plus engageantes	Hebdomadaire
<b>Taux d'ouverture des emails ciblés</b>	% d'e-mails ouverts parmi ceux envoyés avec contenu personnalisé	Augmenter l'efficacité des campagnes (objectif : +15 %)	Par campagne
<b>ROI marketing</b>	Retour sur investissement des actions marketing (CA généré / coût)	+15 % via un ciblage plus pertinent	Mensuel
<b>Répartition par segments clients</b>	Analyse des ventes par segment RFM ou cluster K-Means	Mieux piloter les actions spécifiques à chaque profil	Mensuel
<b>Taux d'erreur de prévision (MAE / RMSE)</b>	Écart entre prévisions de ventes et ventes réelles	Minimiser les erreurs pour optimiser stock & planning	Après chaque cycle de prévision

La mise en place de ces indicateurs et d'un plan de suivi rigoureux garantit une mesure précise de l'impact de la solution IA sur les ventes, l'engagement client et l'efficacité des campagnes marketing. Cette approche orientée résultats permet non seulement de valoriser les apports de l'IA auprès des décideurs, mais aussi d'ajuster dynamiquement les leviers d'action selon les données observées, dans une logique d'optimisation continue.



# Étude des outils et technologies retenus

La sélection des outils constitue un levier stratégique essentiel dans la réussite du projet IA. Elle a été guidée par quatre critères principaux :

- la compatibilité avec les données disponibles,
- la pertinence vis-à-vis des objectifs du projet (recommandation, segmentation, prédiction),
- la facilité d'intégration dans l'environnement technique existant,
- la capacité à industrialiser la solution en cas de généralisation.

Nous avons structuré cette étude autour des différentes phases du cycle de vie du projet IA : acquisition et traitement des données, modélisation, visualisation, déploiement, et monitoring.

## • PRÉPARATION ET TRAITEMENT DES DONNÉES

### Python (langage principal)

- Pourquoi ce choix ?

Python est aujourd'hui le standard dans le domaine de la data science. Il propose une large écosystème de bibliothèques scientifiques et une grande communauté active, ce qui facilite le développement, le partage et la reproductibilité des analyses.

- Alternatives écartées : R (très bon pour la statistique, mais moins adapté aux workflows de machine learning à grande échelle), Java (plus performant mais beaucoup moins souple pour le prototypage IA).

### • BIBLIOTHÈQUES PYTHON UTILISÉES :

BIBLIOTHÈQUE	RÔLE	POURQUOI CE CHOIX ?
<b>pandas</b>	Manipulation tabulaire des données	Manipulation puissante de DataFrames, très lisible
<b>numpy</b>	Calculs numériques	Utilisé comme base de nombreuses autres bibliothèques
<b>gzip / csv</b>	Lecture des fichiers compressés	Indispensable au traitement des .csv.gz

## • MODÉLISATION IA

### Scikit-learn

- C'est la bibliothèque la plus complète et stable pour des modèles classiques (K-Means, Random Forest, régressions...), avec des interfaces simples et standardisées.
- Avantage clé : permet de tester rapidement plusieurs algorithmes et pipelines de preprocessing.

### XGBoost

- Très puissant pour les données tabulaires.
- Résistant au surapprentissage, performant en compétition Kaggle.
- Meilleure précision que les modèles linéaires ou les arbres simples.

### Random Forest

- Complémentaire à XGBoost pour comparer les performances.
- Plus simple à entraîner, facile à interpréter.
- Utile en première approche avant tuning approfondi.

### K-Means (clustering)

- Modèle standard pour les segmentations client RFM.
- Très rapide, lisible, actionnable pour les équipes marketing.

### LightFM (recommandation hybride)

- Combine filtrage collaboratif et contenu produit (prix, catégorie...).
- Supporte bien le problème du cold-start.
- Plus simple à intégrer qu'un système de deep learning.

### Modèles non retenus (et pourquoi) :

- Deep learning (RNN, LSTM) : trop lourd pour le cadre du projet, besoin de GPU, difficile à interpréter.
- Autoencoders pour recommandation : moins explicables, peu adaptés aux données disponibles.
- ARIMA/SARIMA : demandent une stationnarité stricte, difficile à appliquer sur des ventes e-commerce irrégulières.



### **Streamlit**

- Crée facilement des dashboards interactifs.
- Intégrable avec les notebooks.
- Idéal pour exposer les résultats à des non-techniciens.

### **Power BI (option possible si infrastructure Microsoft)**

- Ciblé business.
- Permet des connexions aux données en temps réel.

### **Matplotlib / Seaborn**

- Analyse exploratoire des données.
- Visualisation claire des relations, tendances et clusters.

### **FastAPI**

- Framework rapide et léger pour exposer des modèles IA sous forme d'API REST.
- Facilement déployable avec Docker ou via un cloud provider.

### **Docker**

- Crée des environnements reproductibles et isolés.
- Permet de déployer la même version de code dans tous les environnements (local, cloud, production).

### **GitHub / GitHub Actions**

- Versionning du code source, gestion collaborative.
- GitHub Actions permet l'automatisation des tests et du déploiement.

### **Prometheus + Grafana (prévu en cas d'industrialisation)**

- Monitoring des modèles en production (temps de réponse, erreurs, dérive des performances).



## • DÉPLOIEMENT ET INDUSTRIALISATION : VERS UNE SOLUTION PORTABLE ET MAINTENABLE

Le déploiement d'un système IA ne se limite pas à une simple restitution graphique. Il est important de ne pas confondre les rôles des outils : Seaborn est un outil de visualisation statistique intégré au processus exploratoire ou analytique, tandis que Streamlit est une solution de présentation interactive dédiée aux utilisateurs métier. Ces deux outils ne sont pas interchangeables, mais peuvent être complémentaires si bien intégrés dans une interface unique.

Pour garantir la portabilité de l'application dans tout environnement cible (local, cloud, conteneur), il est indispensable d'embarquer toutes les dépendances. Cela passe par l'usage de Docker, qui permet de créer des images reproductibles et isolées, assurant ainsi le même comportement sur tout poste ou serveur.

En complément de Streamlit, plusieurs alternatives de déploiement peuvent être envisagées selon les besoins :

### FastAPI

- pour exposer les modèles IA sous forme d'API REST, compatible avec des systèmes front-end plus complexes.

### Gradio

- solution légère et intuitive pour créer rapidement des interfaces IA interactives, surtout utile en phase de POC.

- **Power BI** : pour une restitution métier intégrée dans l'écosystème entreprise.





# Roadmap du projet

Le projet a été conçu selon une logique de phases progressives, chacune avec des objectifs, des livrables et des validations internes. La méthodologie employée repose sur un cycle de type agile, permettant des itérations courtes et des ajustements fréquents.

PHASE	DURÉE ESTIMÉE	OBJECTIFS PRINCIPAUX	LIVRABLES
<b>1. Cadrage et audit</b>	Sem. 1-2	Comprendre le besoin, auditer la qualité et structure des données	Note de cadrage, rapport d'audit
<b>2. Préparation des données</b>	Sem. 3-4	Nettoyage, agrégation, transformation des logs utilisateurs	Dataset prêt pour modélisation
<b>3. Exploration et choix des modèles</b>	Sem. 5	Comparaison d'algorithmes, sélection des meilleurs candidats	Notebook comparatif, métriques
<b>4. Modélisation IA</b>	Sem. 6-7	Entraînement des modèles : LightFM, KMeans, XGBoost, etc.	Modèles sauvegardés, code reproductible
<b>5. Tests métier et ajustement</b>	Sem. 8-9	Évaluer les résultats avec les objectifs business	Résultats KPI, ajustements, A/B tests
<b>6. Restitution &amp; visualisation</b>	Sem. 10-11	Création de tableaux de bord, explication des recommandations	Dashboard, visualisation interactive
<b>7. Suivi &amp; plan d'amélioration continue</b>	Sem. 12	Plan de monitoring, recommandations pour industrialisation	Plan de maintenance, fiche modèle





## • ROADMAP DU PROJET : UNE PROGRESSION ITÉRATIVE VALIDÉE COLLECTIVEMENT

La feuille de route du projet repose sur une méthodologie agile, structurée en phases successives : cadrage, audit des données, préparation, modélisation, tests métier, et restitution.

Il est toutefois essentiel de rappeler qu'aucune phase n'est figée de manière définitive. Une fois le cadrage et l'audit validés, ceux-ci ne doivent pas être considérés comme intouchables. Si des éléments critiques émergent en cours de projet (ex. : nouvelle source de données, contrainte RGPD non anticipée, ou changement d'objectif marketing), un retour ponctuel sur une phase antérieure doit rester possible, dans une logique de souplesse contrôlée.

Chaque livrable produit à l'issue d'une phase (note de cadrage, rapport d'audit, dataset prêt, résultats des modèles, tableaux de bord, etc.) doit faire l'objet d'une validation formelle, impliquant les parties prenantes suivantes :

- **Chef de projet IA** : garant de la cohérence technique et du respect des jalons.
- **Data scientist** : responsable de la qualité analytique des livrables.
- **Référent métier** (ex. : responsable marketing) : valide l'adéquation aux besoins opérationnels.
- **DPO ou expert RGPD** : valide la conformité réglementaire si des traitements sont mis en production.

Cette approche garantit un **pilotage rigoureux et collaboratif**, tout en préservant la capacité du projet à s'adapter aux réalités métier et techniques.



# Conformité RGPD et gouvernance des données

La gouvernance des données est au cœur de la stratégie IA d'Amazing. Même si les données sont pseudonymisées, l'analyse de comportements utilisateurs impose une vigilance éthique et réglementaire à toutes les étapes.

## • ANALYSE RGPD

PRINCIPE RGPD	APPLICATION DANS LE PROJET
Minimisation	Seules les colonnes strictement nécessaires sont utilisées (event_type, user_id, product_id, etc.)
Pseudonymisation	Les identifiants utilisateurs sont anonymes (aucun email, nom, localisation).
Finalité explicite	Les traitements sont destinés à la personnalisation marketing uniquement.
Transparence	Tous les traitements peuvent être documentés et expliqués.
Droits utilisateurs	Possibilité d'intégrer une interface permettant à l'utilisateur de refuser la personnalisation.

## • ÉTHIQUE & IA

- **Biais algorithmiques** : les modèles sont testés sur différents segments pour détecter d'éventuelles discriminations.
- **Explicabilité** : importance des variables affichée pour tous les modèles via SHAP ou arbres de décision.
- **Contrôle humain** : les équipes marketing conservent le dernier mot sur les campagnes déclenchées.

# Conformité RGPD et gouvernance des données

## • ÉTHIQUE & IA

- Toutes les données utilisées dans le cadre du projet doivent être anonymisées ou pseudonymisées, conformément à l'article 5(1)(c) sur la minimisation des données et l'article 32 sur la sécurité des traitements. Les données sont stockées selon les normes de sécurité prévues par l'article 25 (protection des données dès la conception) et l'article 32 (mesures techniques et organisationnelles appropriées).
- Le client doit pouvoir exercer à tout moment ses droits d'accès (article 15), de rectification (article 16), d'effacement (article 17), et d'opposition (article 21), via une interface simple et transparente. Ces droits sont garantis par le chapitre III du RGPD, qui encadre les droits des personnes concernées.
- Les algorithmes déployés ne doivent pas prendre de décisions discriminatoires, en accord avec les exigences de l'article 22, qui interdit les décisions automatisées produisant des effets juridiques sans intervention humaine, notamment en cas de discrimination fondée sur le sexe, l'origine ou tout autre critère sensible.
- Des mécanismes de contrôle humain et d'explicabilité sont intégrés dans les traitements algorithmiques pour garantir la transparence, en s'appuyant sur des méthodes telles que SHAP ou les arbres de décision interprétables, conformément à l'esprit de l'article 5(1)(a) sur la licéité, loyauté et transparence du traitement.
- Enfin, une formation spécifique de l'équipe marketing aux implications légales et éthiques de l'intelligence artificielle est prévue, afin d'assurer une appropriation responsable des outils mis en place et une bonne compréhension du principe de responsabilité (article 24) et de la documentation des traitements (article 30).

# Perspectives en cas de mise en production

Si le projet IA venait à être déployé de manière industrielle, plusieurs actions devront être prévues pour respecter les obligations réglementaires :

ACTION	DESCRIPTION
<b>Analyse d'impact (PIA)</b>	Évaluation de l'impact potentiel du traitement sur les droits des personnes concernées. Recommandée pour toute IA de personnalisation.
<b>Désignation d'un DPO</b>	Le Délégué à la Protection des Données serait responsable de l'audit, de la documentation et du suivi des traitements.
<b>Registre de traitement IA</b>	Un registre spécifique documentera chaque pipeline (données sources, objectifs, finalités, durée de conservation, destinataires).
<b>Mécanismes de refus utilisateur</b>	Mise à disposition d'une option "opt-out" dans les paramètres de l'espace client.
<b>Explicabilité des décisions IA</b>	Intégration de modules XAI pour expliquer les recommandations aux utilisateurs ou à la CNIL si nécessaire.

## • PERSPECTIVES D'INDUSTRIALISATION RGPD

Dans le cadre d'un déploiement réel, les actions suivantes sont recommandées :

- **Réalisation d'une Analyse d'Impact (PIA).**
- **Mise en place d'un registre de traitement IA.**
- **Désignation formelle d'un DPO référent pour le projet.**
- **Audit régulier des performances et des biais des modèles en production.**

## • CONCLUSION RGPD

La gouvernance des données dans le projet Amazing repose sur une démarche proactive de conformité. En intégrant des mécanismes de minimisation, pseudonymisation, explicabilité et documentation, le projet respecte pleinement les principes fondamentaux du RGPD.

La conception “Privacy by Design” est au cœur du dispositif, et les actions complémentaires prévues en cas de déploiement assurent une mise en œuvre responsable, éthique et durable de la solution IA.

# Évaluation des moyens prévisionnels et estimation budgétaire

Afin de garantir la faisabilité opérationnelle du projet IA proposé à Amazing, il est indispensable d'évaluer les moyens humains, techniques et financiers à mobiliser. Cette estimation repose sur un déploiement de type preuve de concept (POC), sur une durée estimée de 12 semaines, avec l'objectif de tester, valider, puis itérer la solution avant une éventuelle industrialisation.

## • MOYENS HUMAINS À MOBILISER

Le projet nécessitera une collaboration multidisciplinaire impliquant :

- **1 chef de projet IA** : supervision, planification, coordination avec les équipes métier.
- **1 data scientist** : responsable de l'analyse des données, du choix et de l'entraînement des modèles.
- **1 data engineer (ressource interne d'Amazing)** : mise à disposition des données depuis le Data Lake, traitement initial.
- **1 référent marketing** : interface métier, validation des résultats, tests A/B.
- **1 expert RGPD ou DPO** : validation de la conformité des traitements algorithmiques.

## • MOYENS TECHNIQUES

Le projet s'appuie sur des outils principalement open-source et une infrastructure existante :

- **Environnement de développement** : Jupyter Notebook ou Google Colab.
- **Langage de programmation** : Python.
- **Bibliothèques** : scikit-learn, LightFM, XGBoost, Prophet, SHAP.
- **Visualisation** : Streamlit ou Power BI.
- **Déploiement envisagé** : Docker + FastAPI (si industrialisation).
- **Infrastructure** : Data Lake cloud d'Amazing (déjà en place).

## • ESTIMATION BUDGÉTAIRE

Pour être plus réaliste et défendable, l'estimation budgétaire doit être justifiée par la charge de travail, la durée et le niveau de compétence requis, ainsi que par les coûts d'infrastructure, de formation et d'accompagnement réglementaire.

- **Data scientist** : un projet de 12 semaines à raison de 3 jours par semaine représente environ 25 à 30 jours homme, soit un budget estimé entre **10 000 € et 15 000 €** en fonction du niveau d'expertise.
- **Chef de projet / coordinateur IA** : même sur un rythme partiel, sa présence est nécessaire pour assurer la planification, les échanges avec les métiers et la supervision technique. En moyenne, **4 000 € à 6 000 €** doivent être prévus.
- **Infrastructure cloud** (stockage, traitement, API, monitoring) : même avec une solution AWS ou GCP limitée à un usage ponctuel, les coûts peuvent atteindre **1 500 € à 2 500 €** sur 3 mois, en tenant compte des charges réseau, machines virtuelles et monitoring.
- **Formation ou accompagnement métier** : la sensibilisation des équipes marketing à l'IA, aux tableaux de bord, ou aux biais algorithmiques implique des ateliers ou supports, à estimer autour de **1 000 €**.
- **Accompagnement RGPD ou audit éthique** : si un DPO ou juriste externe est mobilisé ponctuellement, un budget de **1 000 € à 2 000 €** peut être justifié.

## • TOTAL PRÉVISIONNEL RÉALISTE

- entre 17 000 € et 25 000 €, en fonction du périmètre exact, des ressources internes déjà disponibles, et du niveau de professionnalisation visé.
- Cette évaluation permet de mieux anticiper les besoins humains et techniques nécessaires à un déploiement contrôlé, tout en évitant de sous-dimensionner un projet qui engage des traitements de données sensibles, des ressources cloud, et des attentes métier concrètes.

## • ESTIMATION BUDGÉTAIRE

POSTE	COÛT ESTIMÉ	DÉTAIL
<b>Data scientist (3 mois)</b>	~13 500 €	Analyse, modélisation, visualisation
<b>Chef de projet (1 mois ETP)</b>	~4 500 €	Coordination, gestion des livrables
<b>Cloud / infrastructure</b>	~1 200 €	Stockage, calcul, hébergement (POC uniquement)
<b>Sensibilisation métier (1 journée)</b>	~800 €	Ateliers d'appropriation des outils IA
<b>Total POC (hors industrialisation)</b>	~20 000 €	Montant prévisionnel complet

## Note de faisabilité synthétique

Cette note vise à fournir une vue exécutive condensée, destinée aux décideurs ou sponsors du projet. Elle reprend les éléments clés permettant de valider la pertinence et la viabilité de la solution.

### • CONTEXTE

- Amazing souhaite répondre à une baisse significative des ventes sur sa gamme "Basics", en particulier sur les produits de divertissement. Dans un environnement digital concurrentiel, l'entreprise s'oriente vers une exploitation plus fine de ses données client, via l'intelligence artificielle, pour personnaliser l'expérience utilisateur et optimiser ses actions marketing.

### • OBJECTIFS DU PROJET

- Déployer une solution IA visant à :
  - Recommander des produits pertinents à chaque utilisateur,
  - Segmenter les clients selon leur comportement d'achat,
  - Anticiper les ventes pour ajuster l'offre et les campagnes.
- Offrir une visualisation claire et exploitable des résultats aux équipes métiers.

## • SOLUTION PROPOSÉE

- **Recommandation personnalisée** : modèle LightFM (hybride).
- **Segmentation client** : analyse RFM + clustering K-Means.
- **Prévision des ventes** : modèles XGBoost, Prophet, Random Forest.
- Restitution des résultats via dashboards interactifs (Streamlit, Power BI).
- Données extraites du Data Lake cloud d'Amazing.

## • ESTIMATION BUDGÉTAIRE

### RISQUE

### ACTION PRÉVUE

#### Adoption difficile côté métier

Ateliers de sensibilisation, dashboards visuels

#### Données incomplètes ou bruitées

Nettoyage, agrégation, robustesse des modèles

#### Biais algorithmiques

Tests multi-segments, interprétabilité (SHAP)

#### Risques RGPD

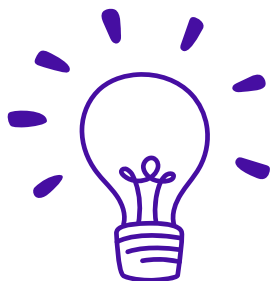
Données pseudonymisées, documentation, supervision DPO

## • VALEURS ATTENDUE

- **+10 à 15 % de taux de conversion estimé via la personnalisation.**
- **Augmentation du panier moyen grâce à des recommandations ciblées.**
- **Meilleure anticipation des ventes pour ajuster le stock et les promotions.**
- **Outil d'aide à la décision basé sur des segments clients concrets.**

## • CONCLUSION

La solution proposée est à la fois réaliste, responsable et à fort potentiel. Elle répond de manière ciblée aux besoins d'Amazing en conciliant performance marketing, compréhension client et conformité réglementaire. Le projet peut être lancé sous forme de POC avec un budget maîtrisé, et monté en puissance selon les résultats obtenus.





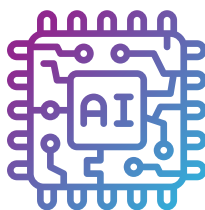
# Conclusion



**Ce projet a permis de démontrer la faisabilité et la pertinence d’une solution d’intelligence artificielle appliquée au secteur du e-commerce, dans un contexte où la performance commerciale dépend de plus en plus de la personnalisation et de l’analyse des données. En répondant à une problématique métier claire – relancer les ventes d’Amazing Basics – nous avons su proposer une approche fondée sur des modèles éprouvés, des outils accessibles, et une logique rigoureuse de cadrage, de test et de validation.**



**La combinaison des algorithmes retenus (LightFM, K-Means, XGBoost, Prophet, Random Forest) permet de couvrir l’ensemble du cycle de décision marketing : de la compréhension du client à la recommandation de produits, jusqu’à la prévision de la demande. Le projet intègre également des dimensions essentielles comme l’éthique algorithmique, la conformité RGPD, la visualisation des résultats, et l’accompagnement des équipes métiers. Ces éléments assurent une solution à la fois performante, responsable, et compréhensible pour tous les acteurs de l’entreprise.**



**En somme, cette mission illustre concrètement comment l’intelligence artificielle peut devenir un levier stratégique au service de la compétitivité et de l’expérience client. La méthodologie suivie, les outils choisis et les livrables proposés offrent une base solide pour un déploiement progressif à plus grande échelle, dans une logique d’amélioration continue. Ce projet est aujourd’hui prêt à passer de la phase d’expérimentation à une mise en œuvre opérationnelle, au service de la transformation digitale d’Amazing.**