

Réalisez le cadrage d'un projet IA

Date de la soutenance : 29/06/2024

Antoine Arragon



Points clés du projet

Contexte :

- L'entreprise Fashion-Insta souhaite développer une application de recommandations vestimentaires basées sur des photos
 - ◆ Les usagers se prennent en photo avec leurs vêtements favoris et obtiennent des propositions d'articles de Fashion-Insta correspondant à leur goût (cf p.3)
- Objectifs :
 - ◆ Augmenter les ventes en permettant l'achat d'articles directement via l'application
 - ◆ Améliorer l'image de marque par une proposition innovante, se démarquer de la concurrence

Dimensionnement du projet (cf p.5-6) :

- Equipe et planning : 8 ETP mobilisés pendant 15 semaines
- Coûts :
 - ◆ Initiaux : développement + infrastructure cloud : 36 245€
 - ◆ Annuels en production : 35 365 €
- Gains estimés : 70 000€ et prévision de croissance de 2%/an par diffusion de l'application
- Rentabilité au bout d'un an et demi selon nos hypothèses

Identification-Gestion des risques et des enjeux légaux et éthiques (cf p.11 à 17)

- L'application nécessite le traitement de données personnelles dites sensibles et requiert une vigilance élevée quant au respect de la réglementation RGPD.
- L'utilisation d'un algorithme de traitements d'images présente le risque de contenir et exposer certains biais, pouvant aller jusqu'à des résultats discriminants. Une attention particulière lors des phases d'entraînement et de test doit permettre d'aboutir à des résultats satisfaisants.
- L'identification de l'ensemble des risques a permis de mettre en place un plan de prévention efficace

Présentation du projet

Fonctionnement de l'application :





Objectifs

1. Financiers :

- Permettre un accroissement des ventes, notamment en ligne
- Prévisions : + 70 000€ de CA la 1^e année puis croissance de 2% ensuite (cf marketing)

2. Image de marque :

- Proposer aux utilisateurs et clients une solution innovante
- Se démarquer de la concurrence

3. Fidéliser la clientèle :

- Capturer un maximum de clientèle et la fidéliser par une amélioration permanente des suggestions de l'application

Ressources mobilisées

Ressources humaines :

8 ETP dont 1 data scientist expérimenté **sous-traitant**

Service / Profil	Junior	Expérimenté	Total
Développeurs backend	2	1	3
Développeurs frontend	1	1	2
Data-scientists	1	1	2
Gestion de projet	0	1	1
Total (ETP)	4	4	8

Infrastructure cloud - Azure

Développement :

- Stockage
- Machines virtuelles optimisées pour le calcul



Production :

- Stockage
- Azure Machine Learning
- Base de données - Azure SQL Database



Ressources financières :

Coûts de développement :

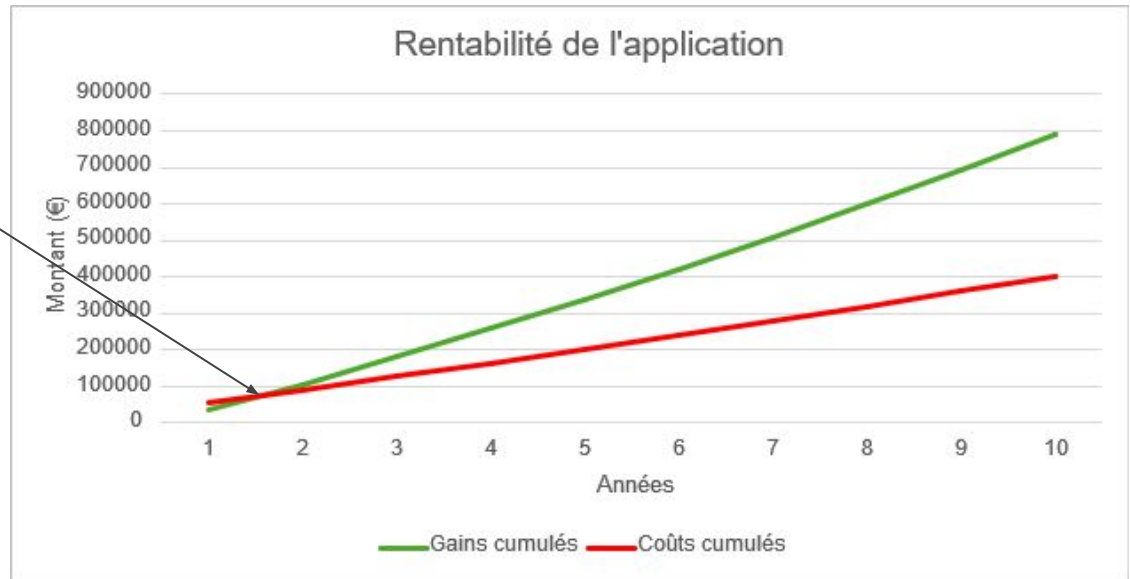
- RH : 31 806 €
- Infrastructure : 4 439 €
- **Total : 36 245 €**

Coûts annuels en production :

- Infrastructure : 29 928 €
- Maintenance : 5 437 €
- **Total : 35 365 €**

Prévision de rentabilité

Rentabilité au bout d'un an et demi :



Hypothèses :

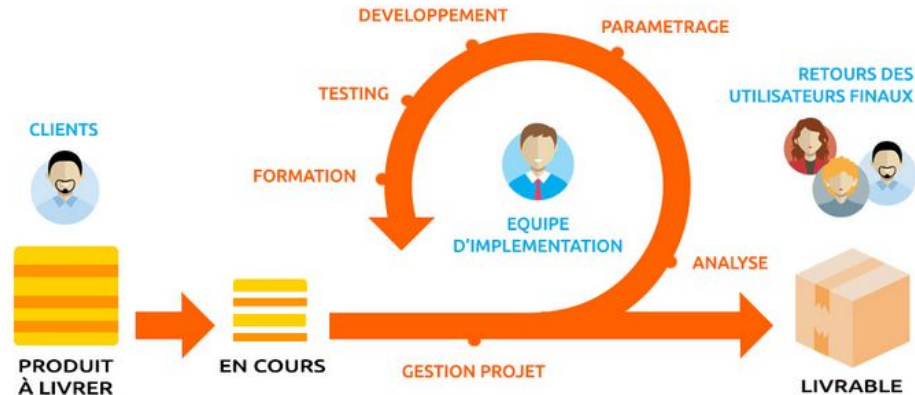
- 6 mois de développement + tests > lancement à la moitié de la 1^{re} année
- Gains attendus : 70 000€ > +2%/an via diffusion croissante de l'application (cf marketing) > inflation estimée à 2%

Gestion de projet

La méthode Agile : méthodologie de gestion plaçant le client au coeur du processus et impliquant l'ensemble des parties prenantes d'un projet en favorisant la communication.

Points clés :

- dialogue entre individus
- division du projet en petites tâches
- souplesse d'organisation (sprints)
- points réguliers sur avancées et/ou blocage
- prise en compte des retours

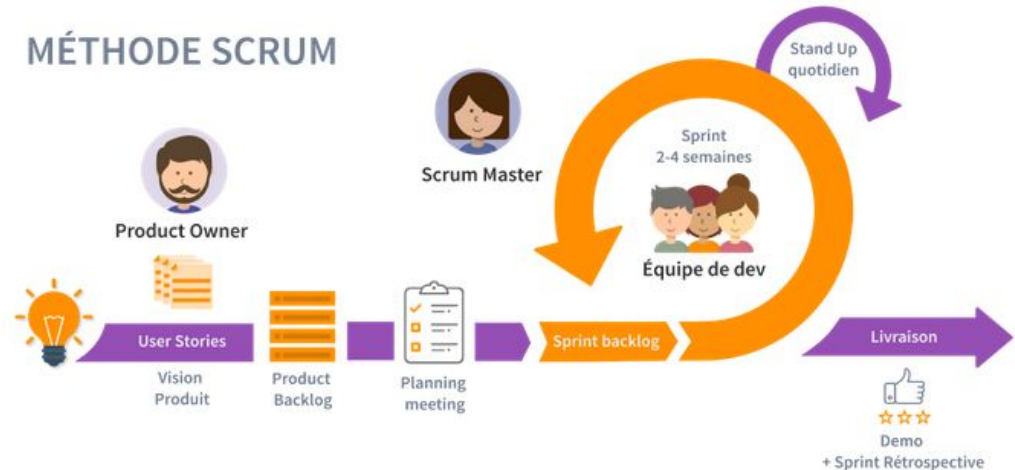


Source image : <https://www.lehibou.com/communaute/methode-agile>

Gestion de projet

La méthodologie SCRUM : adaptabilité quant à la direction du projet en fonction des feedbacks utilisateurs

- 1) Product Backlog
 - a) fonctions à forte valeur
 - b) Priorisation
- 2) Daily Scrum
- 3) Sprints :
 - a) Planning
 - b) Sprint Review
 - c) Sprint Retrospective
- 4) Prise en compte des feedbacks : adaptation des fonctionnalités



Source image : <https://www.lehibou.com/communaute/methode-agile>



Organisation et contenu des points de suivis

Extrait du backlog (voir annexe)

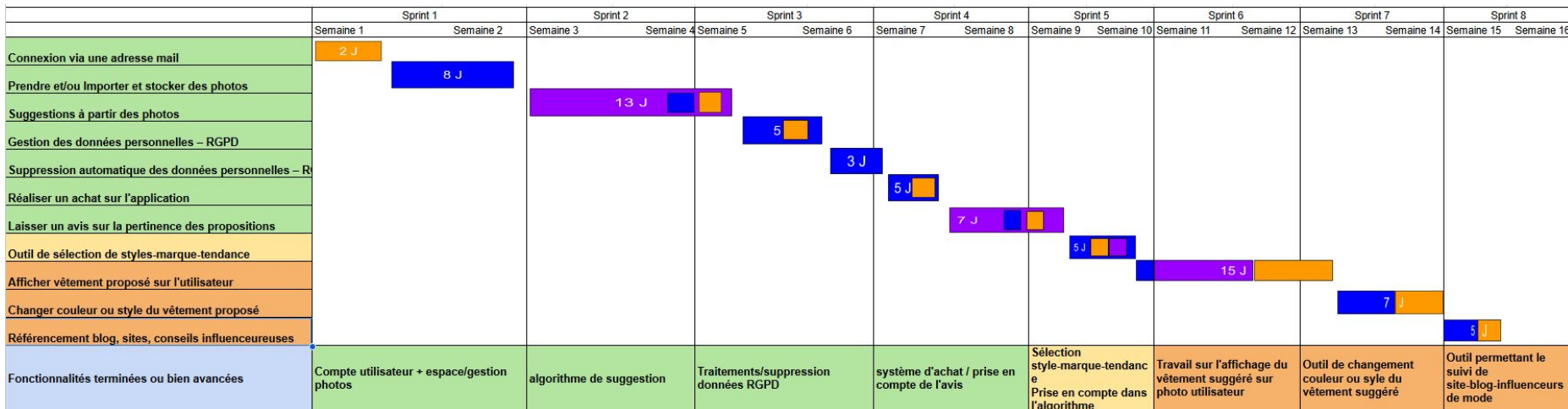
Fonctionnalités	Pondération (1 : très simple – 100 : très complexe)	Méthode Moscow	Priorisation
Connexion via une adresse mail	3	<u>Must have</u>	1
Prendre et/ou Importer et stocker des photos	20	<u>Must have</u>	2
Suggestions à partir des photos	50	<u>Must have</u>	5
Gestion des données personnelles – RGPD	10	<u>Must have</u>	4
Suppression automatique des données personnelles – RGPD	5	<u>Must have</u>	3
Réaliser un achat sur l'application	20	<u>Must have</u>	6
Laisser un avis sur la pertinence des propositions	40	<u>Must have</u>	7
Outil de sélection de styles-marque-tendance	30	<u>Should have</u>	8
Afficher vêtement proposé sur l'utilisateur	100	<u>Could have</u>	9
Changer couleur ou style du vêtement proposé	20	<u>Could have</u>	10
Référencement blog, sites, conseils influenceuses	20	<u>Could have</u>	11

- 11 fonctionnalités
- 8 sprints de 2 semaines
(15 semaines au total)
- 75 Daily Scrum
- 8 Sprints Review
- 8 Sprints Retrospective



Planning des sprints

Développeurs backend
Développeurs frontend
Data-scientists





Enjeux légaux et éthiques

Ce type de projet est soumis à des obligations en matières de protection, gestion et traitements des données personnelles.

RGPD : Règlement général sur la protection des données

➤ encadre et assure la protection des données personnelles au niveau européen.

CNIL : Commission Nationale de l'Informatique et des Libertés

➤ veille au respect de ces règles sur le territoire Français

Champ territorial : Comme il s'agit d'un règlement européen, tous les États membres appliquent le même texte

➤ applicables à tous les acteurs sur le territoire de l'Union européenne

➤ applicables aux opérateurs installés hors de l'Union européenne dès lors qu'ils offrent des biens et des services aux citoyens européens

Champ matériel : s'applique aux traitements de données à caractère personnel

➤ Données personnelles : toute information se rapportant à une personne physique identifiée ou identifiable directement ou indirectement

➤ Traitement : toute opération (procédé automatisé ou non) effectué sur des données personnelles



Enjeux légaux et éthiques

Principes du RGPD :

- Licéité, loyauté et transparence
- Limitation des finalités
- Minimisation des données
- Exactitude
- Limitation de la conservation
- Intégrité et confidentialité des données
- Responsabilité

Logique de contrôle a posteriori ➤ on doit mettre en place des procédures de vérification de conformité

- Obligation d'information
 - fournir une **information extensive** sur le traitement des données aux personnes concernées
 - organiser un **droit d'accès** à ces informations
 - organiser un droit de rectification et d'effacement (droit à l'oubli)
- Mise en place d'un registre des activités de traitement
- Désignation d'un délégué à la protection des données (DPO)
- Sécurisation des données



Enjeux légaux et éthiques

Le registre des activités de traitements

Le registre des activités de traitement permet de recenser les traitements de données et de disposer d'une vue d'ensemble de ce qui est réalisé avec les données personnelles.

Données personnelles	Source	Utilité	Statut
Adresse eMail	Utilisateur	Création d'un compte et accès aux services de l'application	Nécessaire
Photos usagers	Utilisateur	Element à partir duquel les recommandations seront effectuées	Nécessaire
Coordonnées postales	Utilisateur	Envoyer le ou les articles commandés	Nécessaire si achat
Coordonnées bancaires	Utilisateur	Acheter un ou plusieurs articles	Nécessaire si achat
Goûts vestimentaires	Utilisateur	Affiner le profil client	Utile mais pas nécessaire
Age	Utilisateur	Affiner le profil client	Utile mais pas nécessaire
Genre	Utilisateur	Affiner le profil client	Utile mais pas nécessaire
Nom - Prénom	Utilisateur	Personnalisation de l'interface de l'application	Pas nécessaire Doute : Potentiellement nécessaire à l'achat ?

Nom du traitement	Données sensibles ?
Création d'un compte utilisateur	Non
Utilisation d'une photo utilisateur pour suggérer un vêtement	Oui
Achat et livraison d'article(s)	Non
Sélection de préférences clients	Non
Outil de visualisation de l'article recommandé sur photo du client	Oui
Avis du client sur la recommandation	Non



Enjeux légaux et éthiques

L'utilisation de données sensibles

Notre projet d'application nécessite le recueil de photos des usagers. Ces dernières peuvent révéler des données dites sensibles telles que des convictions religieuses ou une orientation sexuelle.

La CNIL nous indique : le règlement européen interdit de recueillir ou d'utiliser ces données, **sauf**, notamment, dans les cas suivants :

- si la personne concernée a donné son **consentement exprès** (démarche active, explicite et de préférence écrite, qui doit être libre, spécifique, et informée) ;
- si les informations sont manifestement rendues publiques par la personne concernée ;
- si elles sont nécessaires à la sauvegarde de la vie humaine ;
- si leur utilisation est justifiée par l'intérêt public et autorisé par la CNIL ;
- si elles concernent les membres ou adhérents d'une association ou d'une organisation politique, religieuse, philosophique, politique ou syndicale.



Enjeux légaux et éthiques

L'identification des biais potentiels du modèle

Les modèles utilisant l'IA peuvent potentiellement engendrer ou conforter différents types de biais : racistes et sexistes par exemple.

En amont :

- contrôler la qualité des données d'entraînement, s'assurer notamment de leur représentativité
- effectuer un prétraitement : floutage, anonymisation des fichiers

Pendant l'entraînement :

- ajuster les paramètres du modèle de manière à réduire les biais
- régularisation
- approches de "bias mitigation"

Après l'entraînement :

- Effectuer des analyses de features importance
- retravailler les données en cas de résultats non satisfaisants
- Change type de modèle ?



Identification des principaux risques

Matrice d'évaluation des risques :						
Probabilité	3				1. Fuite de données sensibles 2. Défaillance du sous-traitant	
	2		7	5 6	1	3. Non conformité RGPD 4. Biais/Discriminations algorithme
	1			2	3 4	5. Infrastructure Cloud sous-estimée 6. Sortie de l'application concurrente
	0					7. Sursollicitation équipes devs
		0	1	2	3	
		Impact				



Plan d'action de mitigation des principaux risques

Risque (événement redouté)	Conséquences (en coût, délai, qualité, satisfaction client)	Impact (0 à 3)	Probabilité (0 à 3)	Criticité (impact*prob)	Actions de prévention (pour éviter l'événement redouté)	Action de correction (si événement redouté avéré)	Priorisation
Fuite de données sensibles	Amende Dégradation de l'image de marque Perte de clients	3	2	6	Limiter la prise de photos aux vêtements seuls Réduction des fonctionnalités (visualisation sur client) Chiffrer les images Sécuriser les accès au stockage et traitements Azure Cloud	Notification à la CNIL Notifiaction aux personnes concernées Audit interne de sécurité	1
Défaillance du sous-traitant (erreur de jugement sur capacités ou erreur d'interprétation et communication par exemple)	Les performances de l'algorithme de suggestion s'en trouverait dégradées et donc l'application elle-même rencontrerait moins de succès et risquerait même de ternir l'image de l'entreprise.	2	1	2	Communication claire des attendus du projet avec le sous-traitant. Garanties contractuelles en cas de "défaut" ou non conformité avec les attendus.	Report du lancement de l'application Préparation d'un plan de communication pour le marketing Embauche d'un data scientist confirmé en interne.	2

Annexes



Backlog

Titre	En tant que	Je veux...	...afin de...	Tâche (côté conception)	Pondération (1 : très simple – 100 : très complexe)	Données nécessaires	Méthode Moscow	Priorisation	Si l'application contient	Si l'application ne contient pas
Connexion via une adresse mail	Utilisateur de l'application	pouvoir me connecter à mon compte sur l'application mobile en utilisant une adresse mail	voir les informations me concernant (infos personnelles, photos prises, articles suggérés, ...)	Création d'un compte utilisateur lié à un couple mail/MDP unique	3	- adresse mail utilisateur - mot de passe	Must have	1	Ce serait un minimum	Cela me dérangerait
Prendre et/ou importer et stocker des photos	Utilisateur de l'application	pouvoir se prendre en photo et/ou importer, stocker et gérer des photos dans l'application	constituer la base de vêtements sur laquelle l'appli devra s'appuyer pour ses propositions.	Création et prévision d'un espace de stockage où l'utilisateur pourra uploader des photos prises précédemment ou prise via l'appareil photo du tel	20	- lien/jacots avec appareil photo du téléphone - accès aux photos stockées sur le téléphone	Must have	2	Ce serait un minimum	Cela me dérangerait
Suggestions à partir des photos	Utilisateur de l'application	que des vêtements me soient suggérés à partir de la ou des photos que je soumetts à l'application	visualiser et potentiellement acheter de nouveaux vêtements	Création d'un algorithme de recommandations d'articles à partir des photos de l'utilisateur	50	- stock conséquent d'images de vêtements labellisées en catégories	Must have	5	Ce serait un minimum	Cela me dérangerait
Gestion des données personnelles – RGPD	Utilisateur de l'application	avoir accès à la gestion de mes données personnelles	de les modifier et/ou supprimer si je le souhaite	Création d'un formulaire de gestion des données personnelles Lien avec les données clients	10	- données personnelles utilisateurs	Must have	4	Ce serait un minimum	Cela me dérangerait
Suppression automatique des données personnelles – RGPD	Utilisateur de l'application	que mes données personnelles soient automatiquement supprimées après un certain temps d'inactivité	que mes droits soient garantis	Programmer une suppression auto des données utilisateurs après un délai fixé.	5	- données personnelles utilisateurs - temps d'inactivité utilisateurs	Must have	3	Ce serait un minimum	Cela me dérangerait
Réaliser un achat sur l'application	Utilisateur de l'application	pouvoir mettre les produits dans un panier, valider une commande, choisir un type de livraison, et terminer par une transaction Bancaire.	d'acheter les vêtements choisis	Création d'un système complet d'achat d'articles avec paiement sécurisé	20	- services de livraison - informations et gestions des infos bancaires - coordonnées et adresse	Must have	6	Cela me ferait plaisir	Je l'accepterais
Laisser un avis sur la pertinence des propositions	Utilisateur de l'application	émettre un avis sur les suggestions de l'appli	d'améliorer ces suggestions, de les faire correspondre de plus en plus aux goûts de l'utilisateur	Prendre en compte ces retour pour affiner et adapter l'algorithme aux préférences utilisateurs	40	- stock d'image de vêtements	Must have	7	Cela me ferait plaisir	Je l'accepterais
Outil de sélection de styles-marque-tendance	Utilisateur de l'application	personnaliser mon profil en indiquant des préférences de style-marque-tendance vestimentaire	d'avoir des suggestions correspondants à mes préférences	- Créer formulaire de préférences - Implémenter ces préférences comme variables de l'algorithme	30	- liste labels/catégories de vêtements - liste actualisée tendance/meilleures ventes ? - liste des marques disponibles	Should have	8	Cela me ferait plaisir	Je l'accepterais
Afficher vêtement proposé sur l'utilisateur	Utilisateur de l'application	pouvoir simuler l'essai d'un vêtement proposé	de voir si ça me va avant de procéder à un achat éventuel	Création d'un système de visualisation de l'article recommandé sur l'utilisateur	100	- photos utilisateur - stock d'image de vêtements	Could have	9	Je l'accepterais	Je n'ai pas d'avis
Changer couleur ou style du vêtement proposé	Utilisateur de l'application	pouvoir apporter des modifications au vêtement suggérer (style et/ou couleur)	de voir si ça correspondrait à mes goûts.	Proposer des articles similaires à celui recommandé par l'algorithme (couleur, formes proches)	20	- photos utilisateur - stock d'image de vêtements	Could have	10	Je l'accepterais	Je l'accepterais
Référencement blog, sites, conseils influenceuses	Utilisateur de l'application	Avoir accès à des blog, sites, conseils influenceuses	guider mes choix	- Créer un recueil de sites de référence mode - Se servir des éléments les + visualisés par l'utilisateur pour affiner ses préférences et améliorer les recommandations	20	- liste de sites mode/influenceuse s/blogs	Could have	11	Je n'ai pas d'avis	Je l'accepterais

Plan d'action de mitigation des principaux risques (complet)

Facteurs de risque	Risque (événement redouté)	Conséquences (en coût, délai, qualité, satisfaction client)	Impact (0 à 5)	Probabilité (0 à 5)	Criticité (Impact*prob)	Actions de prévention (pour éviter l'événement redouté)	Action de correction (si événement redouté avéré)	Priorisation
Etant donné que ...	si ...	Alors ...						
Les développeurs de l'application mobile travaillent en parallèle sur une autre application urgente	un problème advient durant le développement de cette autre application	Cela impacterait le développement de notre application, on pourrait prendre du retard et/ou perdre en qualité ou en nombre de services proposés.	1	2	2	2. Communication régulière avec les responsables du développement de l'autre application. Adaptation/modification des sprints si besoin.	Supprimer ou reporter certains services de l'application en fonction de la priorisation effectuée dans le backlog	7
Le data scientist est junior et que les modèles sont développés par un sous-traitant	Défaillance du sous-traitant (erreur de jugement sur capacités ou erreur d'interprétation et communication par exemple)	Les performances de l'algorithme de suggestion s'en trouveraient dégradées et donc l'application elle-même rencontrerait moins de succès et risquerait même de ternir l'image de l'entreprise.	2	1	2	2. Communication claire des attentes du projet avec le sous-traitant. Garanties contractuelles en cas de "défaillance" ou non conformité avec les attentes.	Report du lancement de l'application Préparation d'un plan de communication pour le marketing Embauche d'un data scientist confirmé en interne.	2
Les délais sont très courts car un concurrent planifie le même type d'application, il faut être les premiers à proposer ces services	Le concurrent sort le même type d'application avant nous	On devrait revoir à la baisse les gains estimés par la sortie de l'application. Risque de perte de clients et de baisse de satisfaction sur notre marque.	2	2	4	4. Etablissement d'un planning de sprints prenant en compte ces éléments tout en restant réalisable. Possibilité de le modifier si des informations inquiétantes sur l'avancée du concurrent parviennent. Possibilité de sortir l'application sans y avoir inclus tous les services.	Etudier l'application concurrente et se démarquer par l'ajout d'une fonctionnalité innovante. Y réfléchir pendant le développement de l'application et lors des réunions sprints. Prévoir une campagne de marketing agressive pour retenir ou capter les clients.	6
Les données de type image sont des données personnelles sensibles	Fuite de données sensibles	Amonde Dégradation de l'image de marque Perte de clients	3	2	6	6. Limiter la prise de photos aux vêtements seuls Réduction des fonctionnalités (visualisation sur client) Chiffrer les images Sécuriser les accès au stockage et traitements Azure Cloud	Notification à la CNIL Notification aux personnes concernées Audit interne de sécurité	1
La gestion des données personnelles et leur sécurité est un enjeu majeur, pour rassurer les clients de poster leurs photos et leurs préférences	Non conformité au RGPD	Amonde Dégradation de l'image de marque Perte de clients	3	1	3	3. Inclure le DPO comme partie prenante du projet dès son initialisation Faire un point sur la réglementation RGPD et la conformité des traitements effectués à chaque sprint. Documenter et mettre à jour le registre des traitements à chaque sprint.	Notification à la CNIL Notification aux personnes concernées Formations internes aux réglementations RGPD	3
Mauvaise qualité du jeu de données d'images initiales servant à l'entraînement du modèle	Algorithme non performant Recommandations peu fiables Biais et discriminations potentielles	Dégradation de l'image de marque Perte de clients	3	1	3	3. Vigilance quant à la sélection des images constituant la base sur laquelle le modèle sera entraîné. Vérification constante des prédictions du modèles lors de la phase de test de l'application. Application d'une régularisation et différents prétraitements de données afin de limiter les biais au maximum.	Mise en maintenance de l'application le temps de réentraîner intégralement le modèle. Campagne marketing pour limiter la déception et la perte de clients.	4
Phase d'entraînement du modèle trop limitée	Mauvaise appréciation de l'infrastructure cloud à mettre en place Limitation de l'entraînement du modèle.	Coûts supplémentaires engendrés par l'acroissement des capacités à mettre en place pour avoir un modèle suffisamment performant.	2	2	4	4. Concentration et validation par des experts de l'architecture cloud mise en place Suivre les coûts engendrés par cette infrastructure.	Révision des prix de l'application pour prendre en compte la hausse des coûts. Réflexion stratégique sur la politique commerciale à mettre en place suite à cette hausse de coûts.	5

Merci de votre attention.

