Vendredi 4 Octobre 2024

Présentation

NÉBULARIUM

présenté par Antoine Desrochers-Godin
Cristel Silva
Houssam Bourkane

Preuve de concept

L'objectif de ce projet est de développer une interface utilisateur qui permet de découvrir des livres francophones "invisibles" en utilisant la sérendipité (1) comme principe de navigation. Les livres seront représentés sous forme d’étoiles qui forment une constellation de points reliés entre eux en fonction de leur similitude. Des filtres interactifs permettront d'affiner l'affichage de la constellation.

1 La sérendipité est le fait de faire par hasard une découverte inattendue qui s'avère ensuite fructueuse (wikipedia, Sérendipité)

Présentation du projet

Quoi?

Ce projet a pour objectif de révéler les contenus francophones souvent "invisibilisés" et de favoriser la bibliodiversité en offrant à l'utilisateur la possibilité d'accéder aisément à des documents difficilement accessibles et de s'informer à leur sujet.

Qui ?

Ce projet s'adresse aux usagers d’une bibliothèque, qui n’ont pas d’idée précise de recherche ou qui sont peu familiers avec les moteurs de recherche et/ou avec la langue française. Notre solution cible également les lecteurs plus expérimentés souhaitant découvrir des œuvres inconnues en lien avec leurs intérêts habituels.

Expérience utilisateur

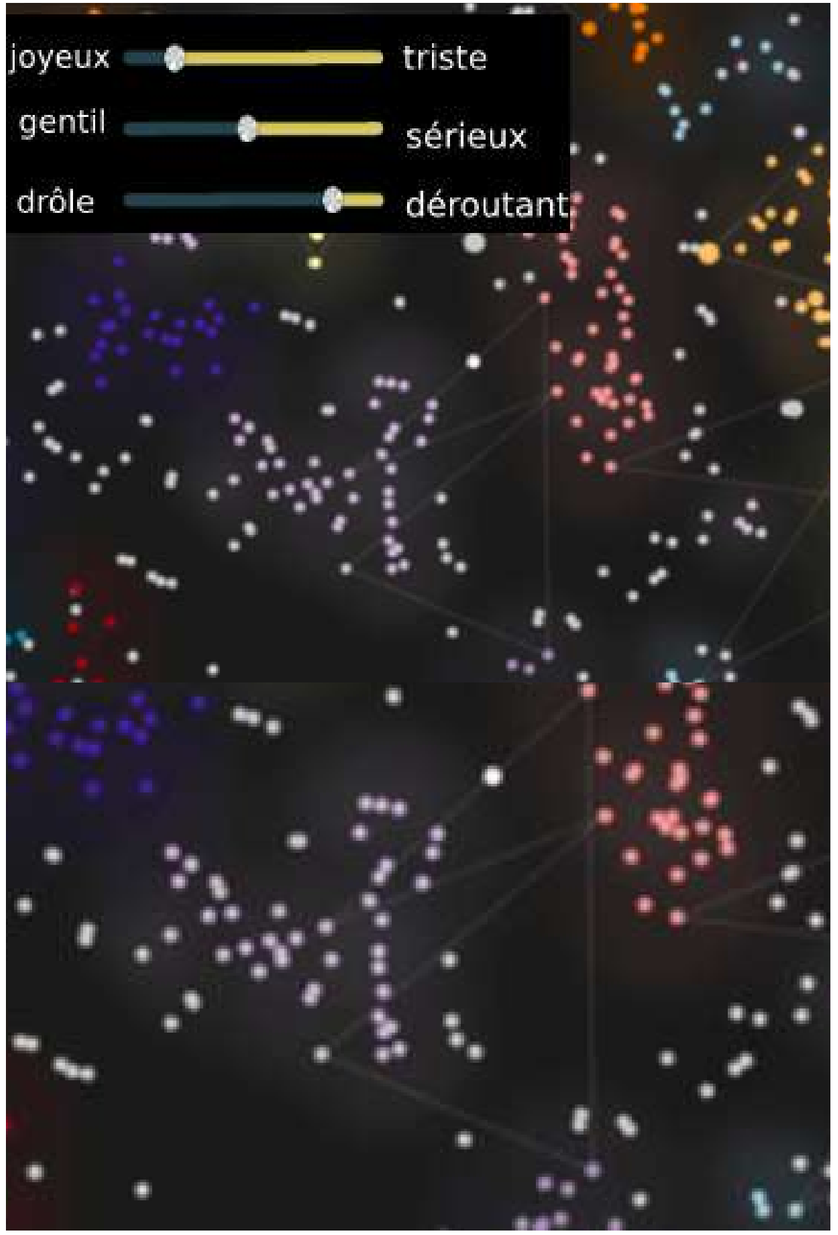
Afin de faciliter l'accès de ces documents aux usagers, nous proposons une alternative au modèle classique de moteur de recherche retrouvé dans un catalogue de bibliothèque.

Et si nous pouvions réaliser une recherche à travers notre humeur ? À travers le degré d'optimisme ou de pessimisme d'un livre ? Son niveau de violence ? C'est le défi que nous nous sommes donnés.

L'interface prend la forme d'une constellation où chaque étoile représente un document. En fonction du choix d'émotion de l'utilisateur, certaines étoiles brillent, grossissent, d'autres disparaissent, deviennent moins lumineuses en suivant l'affinité de l'utilisateur.

Le degré de rapprochement entre les points varie selon la proximité des sujets associés aux documents.

Certaines de ses étoiles brillent également plus que d'autres en fonction de leurs données d'emprunt. De sorte à ce que le l'utilisateur soit porter à explorer autour de sujets populaires.



Présentation du projet

Expérience utilisateur (suite)

Pour préciser sa recherche, l'utilisateur dispose également de filtres afin de l'orienter vers un sujet plus en lien avec ses intentions.

Il peut donc filtrer les étoiles visibles selon:

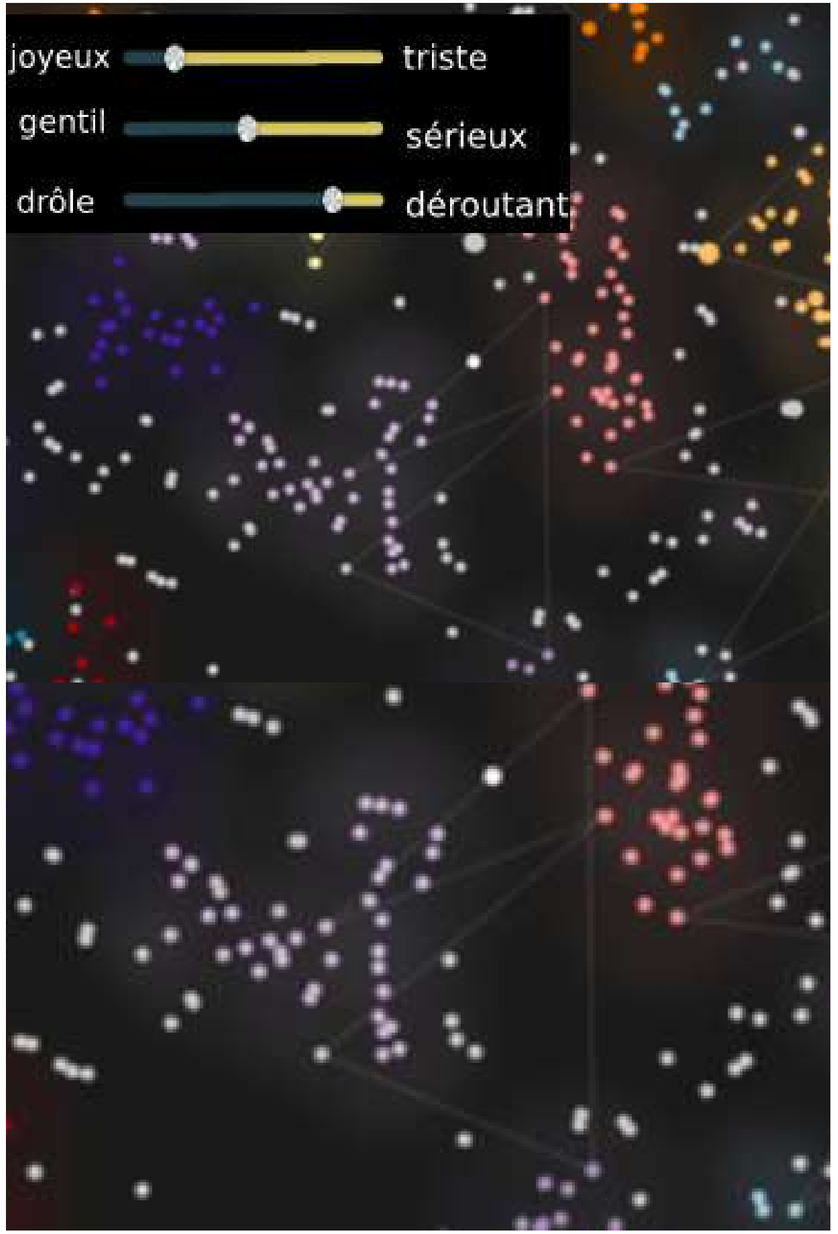
- le pays / province
- le genre littéraire
- la longueur du document (nbr. de pages)
- le public (jeunesse, adulte)
- le type de document

Lui seront proposés des contenus issus de la francophonie organisés selon ses choix.

références / inspirations :

https://www.whichbook.net/mood-emotion/

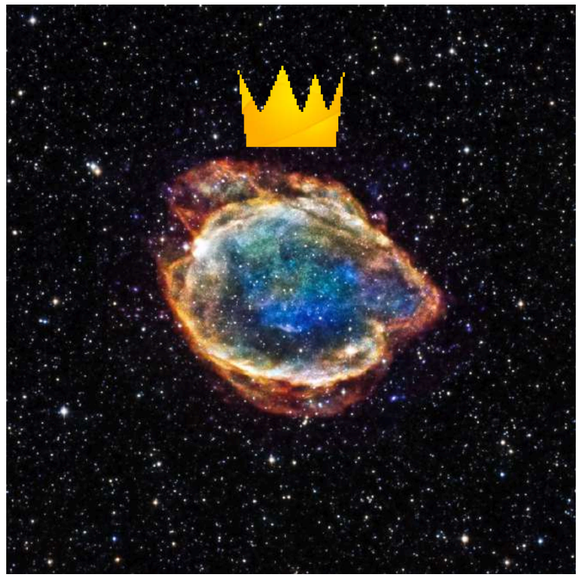
https://audiostellar.xyz/lang/en/index.html



Expérience utilisateur

l’accessibilité

Dans le but de faciliter la compréhension de l’interface (constellation), l’utilisateurl sera guidé par un tutoriel intuitif. Cultiver le mystère demeure toutefois un élément important de l’expérience.



Pour ce faire, nous présenterons les “Étoiles du jour”, choisies par un afin que l’utilisateur puisse comprendre rapidement ce que chaque étoile contient.

Ces étoiles seront déterminées par une donnée de consultation effectué dans les 7 derniers jours.

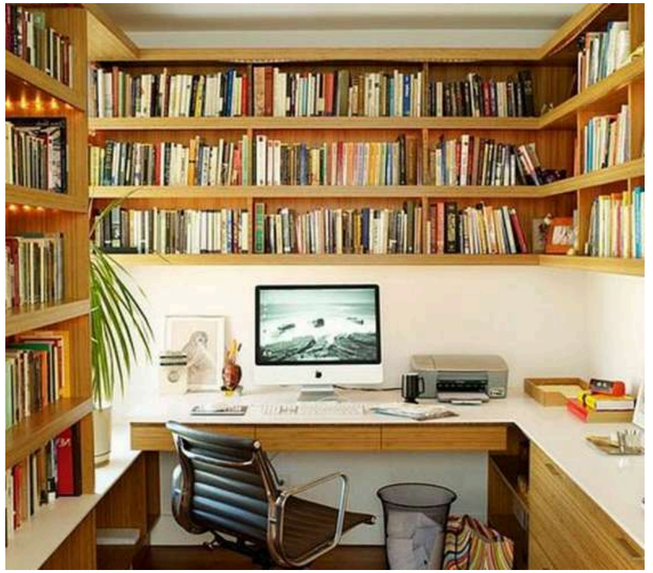
Technologies utilisées

L’interface est *responsive*. Nous pouvons donc l’implémenter comme plugiciel sur les ordinateurs de recherche en bibliothèque, comme interface web pour y avoir accès de chez soi et dans des bornes à grands écrans pour une expérience optimales.

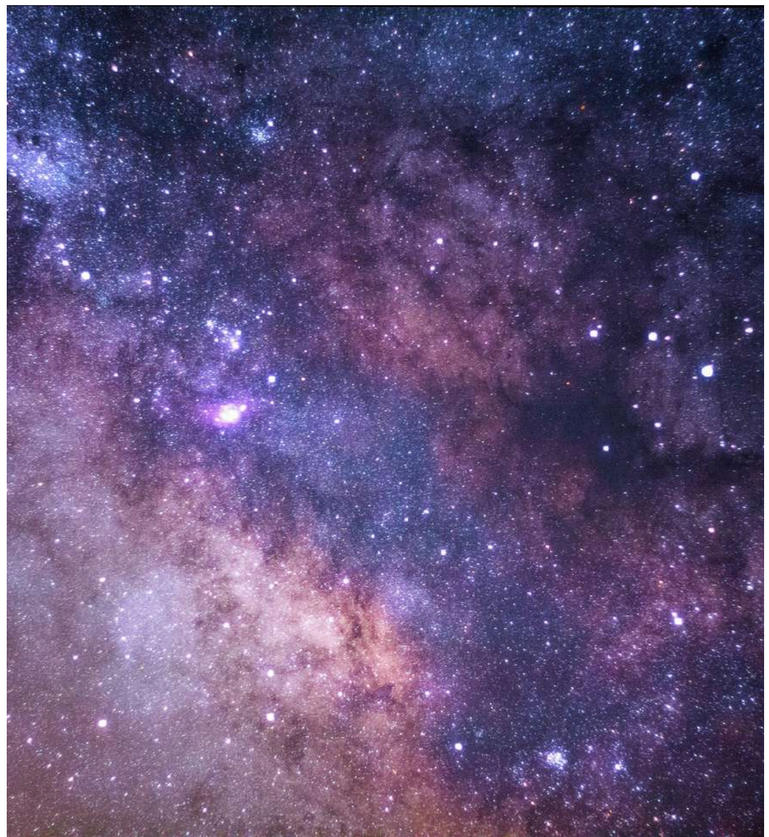
Nous utiliserons l’IA, plus précisément des tecniques de fouille de texte et de machine learning pour déterminer le rapport de proximité entre nos documents.

L’impact attendu

L’aspect ludique et interactif de la solution permettra d’augmenter le degré d’engagement de l’utilisateur lors de sa recherhe. Nous souhaitons promouvoir une “déambulation” dans les étoiles de l même manière dont quelqu’un déambulerait dans une bibliothèque à la recherche d’un document.
Cette expérience pourrait ainsi avoir un impact positif sur les statistiques d’emprunts des livres francophonie portés à l’oubli.



Utilisation des données fournies



Processus d'identification des contenus francophones

Pour évaluer ce qui constitue une ressource francophone, notamment dans le cas des livres, l’addition de trois critères permet de filtrer le contenu francophone retrouvé dans les données:

Pays + Langue + Maison d’édition

Utilisation des données BAnQ pour la création de listes

- Langue: Analyse des jeux de données fournis contenant des champs sur la langue des documents afin d'établir une liste de codes linguistiques ISO 639-1 pour le français (fr) et MARC (fre).
- Pays: Analyse des jeux de données fournis contenant des champs sur les pays d'origine des documenst afin d'établir une liste de codes MARC et ISO 3166 pour les pays dont la langue officielle est le français.
- Maisons d'édition: Analyse des jeux de données fournis contenant des champs sur l'éditeur afin d'établir une liste de maisons d'édition francophones (exemple de liste ici: https://www.chasse-aux-livres.fr/editeurs)

Utilisation des outils d’analyse de texte pour l’enrichissement de données

L’utilisation des techniques et outils d'analyse de sentiment et de texte pour compléter les données existantes sera effectuée sur deux corpus de textes différentes:

1. Corpus comprenant des commentaires de livres afin d’identifier les émotions liées à chaque ouvrage (définition du paramètre “émotion”)

+

2. Corpus comprenant des quatrièmes de couverture des livres afin d’identifier des thèmes et des mots-clés de chaque ouvrage qui ne se trouvent pas dans les notices bibliographiques (définition du paramètre de “rapprochement” entre les termes par la similarité des sujets reliés à un ouvrage).

Étapes prévues:

- Rassemblement des corpus
- Prétraitement des données
- Choix de l'approche d'analyse de sentiment/ de texte
- Analyse et interprentaion des résultats

Utilisation du schéma du modèle IFLA LRM

L’utilisation prévue du schéma du modèle IFLA LRM (Library Reference Model), intégrée par DataBnf, peut considérablement améliorer et enrichir les liens de proximité entre les documents présentés par l’interface (points raprochés dans la constellation) en plus d’inclure autres types de documents (films, audio, images, etc.) non présents dans les jeux de donnés fournis initialement.

Étapes prévues:

- Compréhension et application du modèle IFLA LRM (Oeuvre, Expression, Manifestation, Item).
- Enrichissement des métadonnées et des vocabulaires contrôlés (RVM, Rameau) par l’intégration au modèle LRM à des champs spécifiques.
- Amélioration des suggestions contextuelles et augmentation de la visibilité des contenus par l’utilisation d’algorithmes de recommandation basés sur les relations IFLA LRM.

STACK TECHNIQUE
(TECHNOLOGIES UTILISÉES)

1. Backend
2. Frontend
3. Base de données
4. Visualisation des données
5. Algorithme
6. Infrastructure

BACKEND

Le backend doit permettre de gérer la base de données des livres, de servir les informations via une API, et d'exécuter les algorithmes de recommandation et de filtrage.

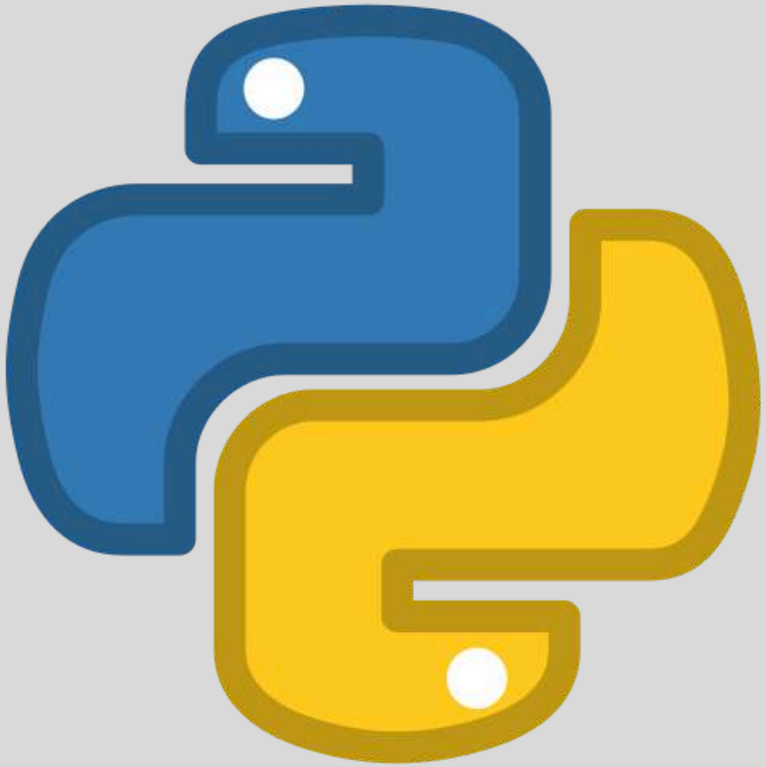
Technologie utilisée : Python

Pourquoi ?

- Python est un langage de choix pour le développement d'applications intégrant des algorithmes de machine learning, et les frameworks Flask ou Django sont idéaux pour construire des API performantes, tout en restant légers et faciles à déployer.

Avantage :

- Intégration fluide avec les bibliothèques de data science : Grâce à Python, il est possible d'intégrer des outils comme scikit-learn, TensorFlow, ou encore pandas pour gérer les aspects complexes de la recommandation de livres.



FRONTEND

Le frontend est la partie visible par les utilisateurs, où l'interface doit être à la fois réactive, interactive et visuellement fluide.

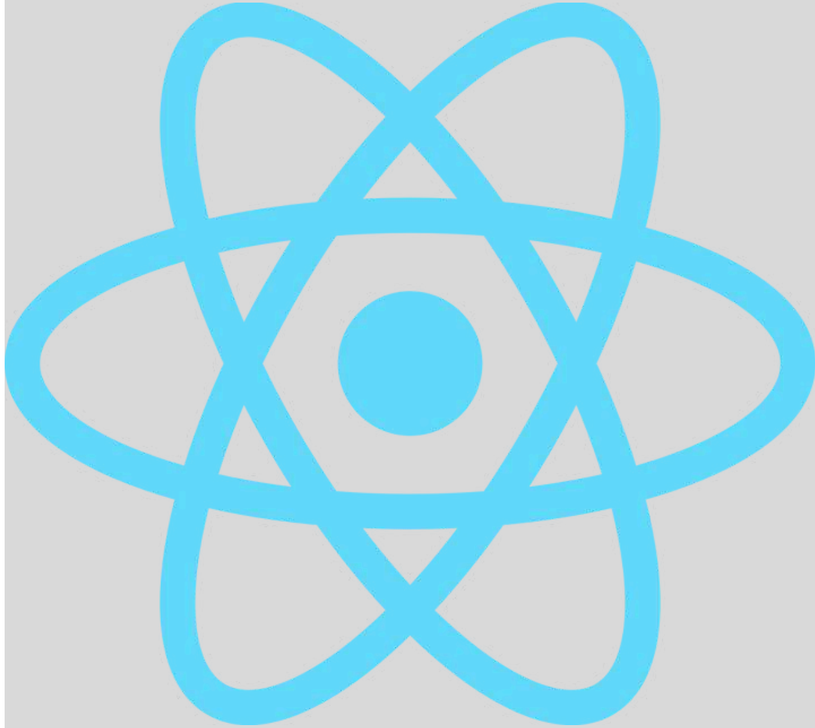
Technologie utilisée : React

Pourquoi ?

- React.js est un framework JavaScript moderne, particulièrement apprécié pour la création d'interfaces utilisateur interactives et réactives. Il est bien adapté aux applications web nécessitant beaucoup d'interactions, comme une bibliothèque virtuelle.

Avantage :

- DOM virtuel performant : La gestion efficace du DOM (Document Object Model) par React permet de maintenir des performances élevées, même lors de manipulations complexes comme le filtrage dynamique ou le zoom sur une carte de points.



BASE DE DONNÉES

La base de données doit stocker toutes les informations sur les livres (métadonnées, relations, etc.) et permettre de créer des liens dynamiques entre ces données pour la visualisation.

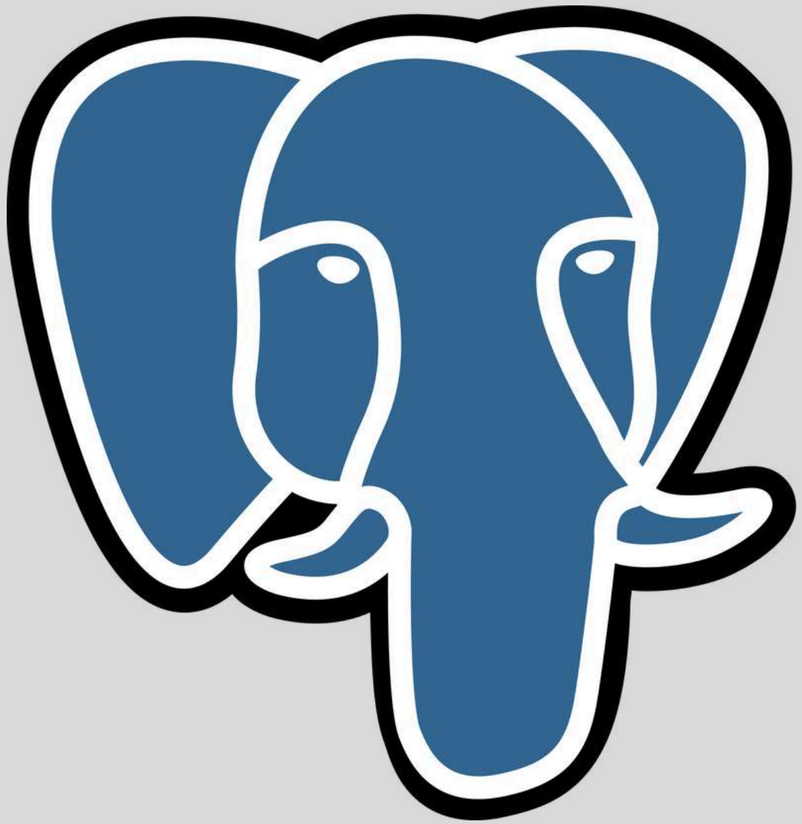
Technologie utilisée : PostgreSQL

Pourquoi ?

- PostgreSQL est une base de données relationnelle robuste, parfaitement adaptée à des structures de données rigides où les relations entre les livres, auteurs, genres, et sujets sont bien définies.

Avantage :

- Gestion avancée des requêtes complexes : PostgreSQL est capable de gérer des requêtes SQL complexes, ce qui est essentiel lorsque l'on souhaite faire des jointures complexes pour trouver des similitudes entre les livres.



VISUALISATION DES DONNÉES

Pour la partie visualisation de la constellation de livres, il est crucial de choisir une librairie qui permet de créer des graphes interactifs, en gérant les animations, le zoom, et la manipulation des points.

Technologie utilisée : D3JS

Pourquoi ?

- D3.js est une bibliothèque JavaScript puissante pour créer des visualisations de données interactives en 2D, qui s'adapte parfaitement à une carte de points montrant la similitude des livres.

Avantage :

- Idéal pour représenter une constellation de points avec des liens, qui peuvent être filtrés et modifiés en fonction de critères spécifiques.

ALGORITHME

Les algorithmes sont au cœur du système de suggestion, permettant de déterminer la similitude entre les livres et de proposer des recommandations pertinentes.

Technologie utilisée : Scikit-learn

Pourquoi ?

- Scikit-learn est l'une des bibliothèques de machine learning les plus populaires en Python. Elle est adaptée pour implémenter des modèles classiques comme K-Nearest Neighbors (KNN), qui est bien adapté pour mesurer les similitudes entre les livres en se basant sur des caractéristiques comme le sujet.

Avantages :

- Avec Scikit-learn, tu peux facilement tester différents algorithmes pour ajuster les propositions et voir ce qui fonctionne le mieux.

INFRASTRUCTURE

Pour garantir la sécurité des données et la scalabilité de l'application, certaines technologies doivent être mises en place

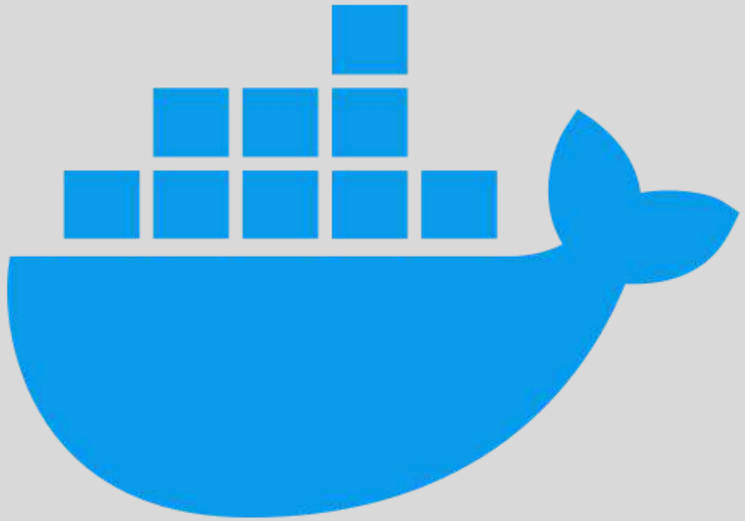
Technologie utilisée : Docker & Kubernetes

Pourquoi ?

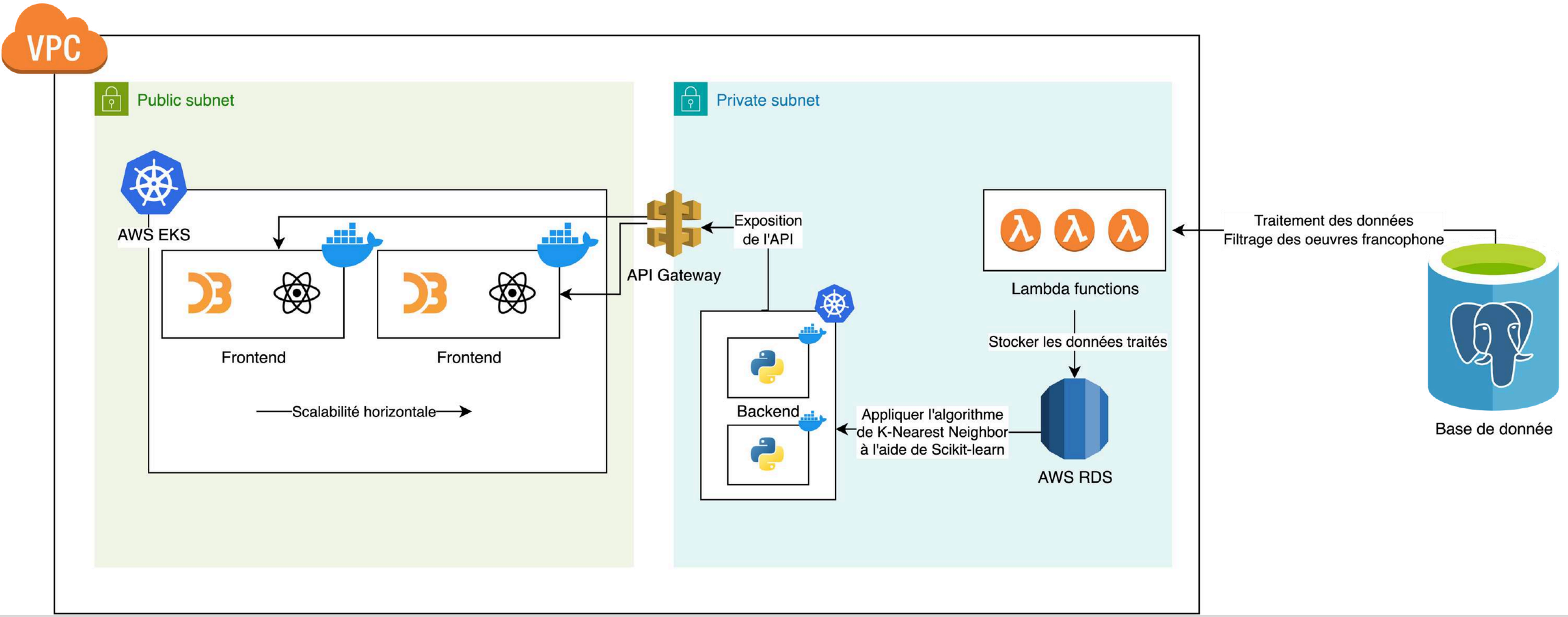
- Docker est une technologie de conteneurisation qui permet d'encapsuler ton application dans des environnements isolés et reproductibles. Kubernetes est un orchestrateur qui permet de déployer et gérer ces conteneurs à grande échelle.

Avantages :

- Scalabilité horizontale : Kubernetes facilite l'ajout ou la suppression de ressources selon la charge, ce qui est essentiel pour une application qui doit gérer de nombreux utilisateurs ou une grande quantité de données.



Architecture de la plateforme



Flux du modèle

K-Nearest Neighbors (KNN) : Calcul des distances entre les livres en fonction de leurs **sujets** et autres paramètres pour trouver les plus proches voisins.

collecter les métadonnées des livres auprès des catalogues en ligne

COLLECTE DE DONNÉES

filtrer la francophonie en se basant sur (code du pays, code de langue et maison d’édition) ou sujet

PRÉPARATION DES DONNÉES

SÉLECTION DE L’ALGORITHME DE SIMILITUDE

ENTRAÎNEMENT DU MODÈLE

Entraîner le modèle pour capturer les relations entre les différents paramètres et identifier les livres similaires grâce au machine learning.

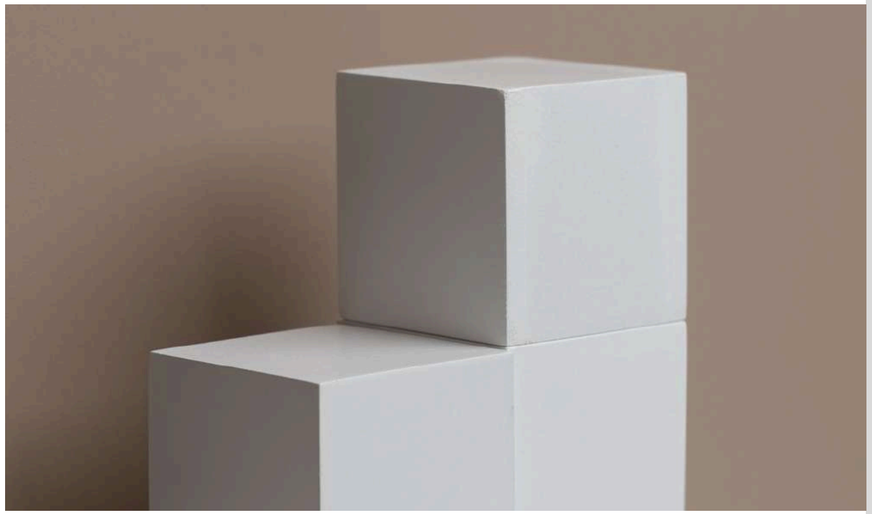
ÉVALUATION DU MODÈLE

Mesurer la performance en vérifiant si les livres proposés comme similaires partagent effectivement des sujets proches et des paramètres communs.

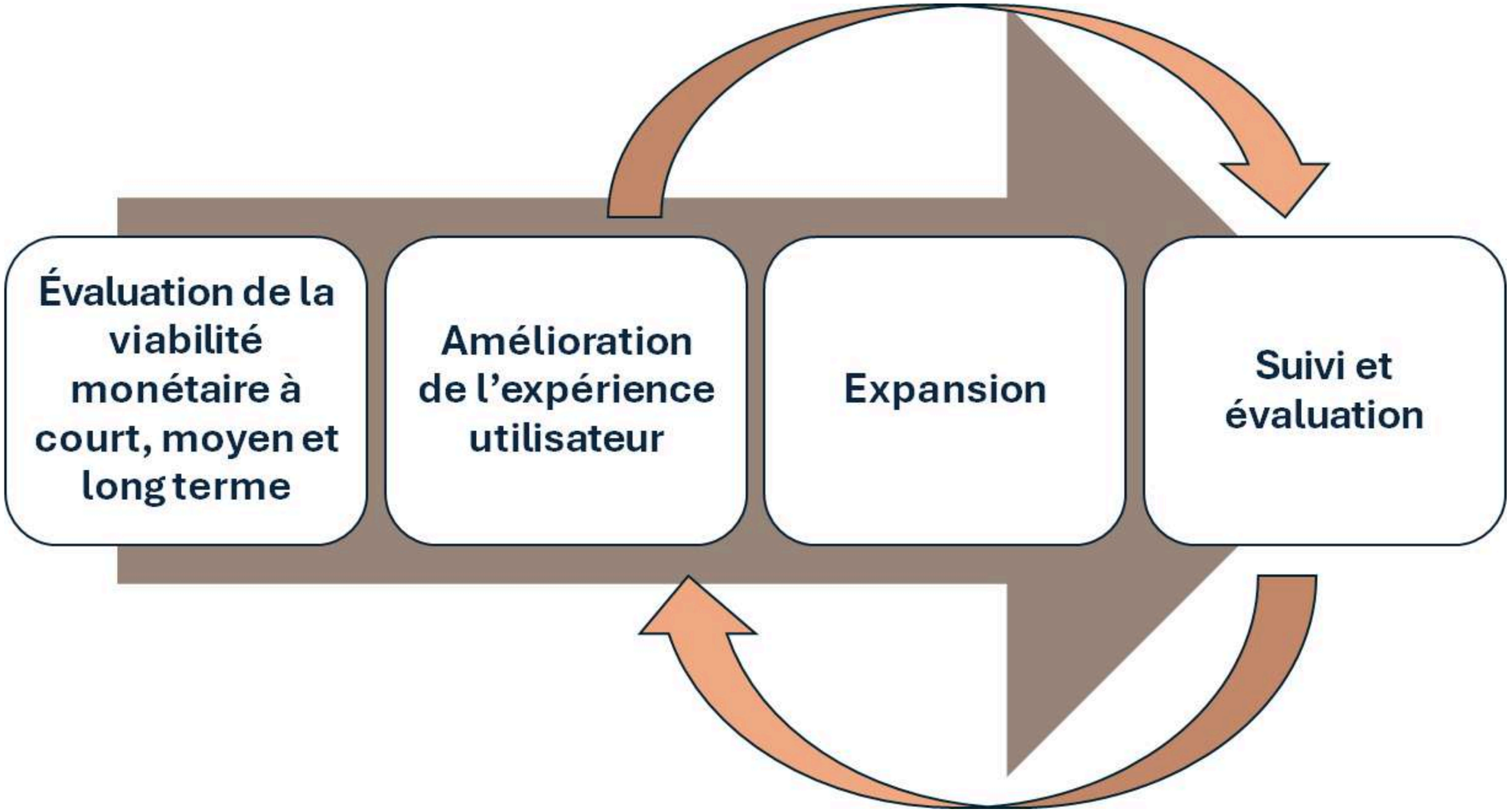
PROPOSITIONS/SUGGESTION

À chaque recherche ou interaction de l'utilisateur, prédire et afficher les livres les plus similaires en utilisant les distances calculées sur la base des sujets et des autres paramètres.

Déploiement



Phases de déploiement



Phases de déploiement- Évaluation de la faisabilité et de la viabilité monétaire à court, moyen et long terme.

- Évaluer les besoins techniques et de financement pour le développement initial de la solution, la socialisation (marketing) et l’amélioration en continu.

- Explorer des options de financement pour soutenir la croissance de la solution. On pense notamment à des subventions et à des programmes de soutien à l'innovation numérique (SODEC, Ministère de la Culture et des Communications, BPI France, startups, etc.)

- Développer des modèles des partenariats pour assurer la viabilité économique.

Phases de déploiement- Amélioration de l’expérience utilisateur.

- Tester une prémière version de la solution (version beta) auprès d’un échantillon de personnes afin de collecter des commentaires.

- Prévoir du temps de formation et d’assistance technique auprès des utilisateurs, et ce, durant tout le cycle de vie de la solution.

- S'assurer que l'expérience utilisateur est fluide, agréable et libre de problèmes grâce aux commentaires (feedback) des utilisateurs.

- Évaluer les métriques (tracking) d’utilisation de la solution afin d’assurer une amélioration en continu.

Phases de déploiement- Expansion

- Identifier de nouveaux marchés, francophones et autres, pour étendre la portée de la solution à partir des contenus francophones traduits.

- Adapter le contenu proposé aux spécificités culturelles de chaque région.

- Utiliser des stratégies pour attirer de nouveaux utilisateurs.

Phases de déploiement- Suivi et évaluation

- Définir des indicateurs de performance clés (KPIs/ tracking), tels que le taux de conversion des recherches, les taux d’emprunt et le taux de croissance des utilisateurs pour mesurer le succès de la mise à l’échelle.

- Mettre en place des outils de collecte des retours d'expérience, d’engagement, de satisfaction des utilisateurs et s’adapter en conséquence.

- Générer des rapports d'analyse pour évaluer l'évolution.

ÉVALUATION DE L’ÉQUIPE

Expérience collective

- Nos compétences diverses se sont bien complémentées de manière général. La répartition des tâches s’est fait sans friction.
- Utilisation de Coworking space, Discord pour échanger en temps réel, partager des idées, et résoudre rapidement les problèmes techniques rencontrés.
- Faiblesse dans la structuration du temps.

Pour le futur..

- Mettre en place des routines de communication plus structurées (meetings hebdomadaires, outils de suivi des tâches) pour suivre efficacement les progrès et résoudre les défis en équipe.
- Encourager la formation continue, chaque membre pouvant explorer de nouveaux domaines techniques ou méthodologiques pour diversifier les compétences globales de l’équipe.
- Organiser des sessions de brainstorming ou de résolution de problèmes en groupe pour favoriser une collaboration plus créative et interactive.
- Envisager l'intégration de nouvelles compétences ou l'élargissement de l’équipe si nécessaire, notamment dans des domaines comme le marketing ou le business development, pour élargir les perspectives du projet.