

# Rapport de projet : Création de mon assistant de portfolio

**Julien Forestier** — BUT Science des Données

## Pourquoi ce projet ?

Dans le cadre de mon BUT, j'ai dû travailler sur les nouveaux outils d'intelligence artificielle. L'idée de ce projet était de créer un assistant capable de répondre aux questions sur mon parcours, mes compétences et mes projets à ma place. Plutôt que de laisser l'IA répondre n'importe quoi, j'ai utilisé une technique appelée RAG. Cela permet de "forcer" l'IA à aller chercher ses informations dans des documents que j'ai moi-même préparés, comme mes fiches de projets ou le détail de ma formation en Science des Données. Le but était d'avoir un outil fiable, qui ne s'invente pas une vie et qui présente mon profil de façon professionnelle aux recruteurs.

## La préparation des données et le découpage

Tout a commencé par la rédaction de mes fichiers sources. J'ai créé plusieurs documents en format Markdown pour détailler mes expériences, mes compétences en programmation (comme le R, Python ou Scala) et mes projets.

```
data > 05_Bilan_Objectifs.md > abc # Bilan et objectifs > abc ## Objectifs pour la troisième année
1   # Bilan et objectifs
2
3   ## Bilan académique
4 Mes résultats sont corrects et suffisants pour valider l'année, mais ils ne sont pas encore pleinement à la hauteur de mes ambitions. Ils ne me placent pas encore dans les meilleures conditions pour intégrer un master sélectif, ce qui m'a amené à prendre du recul sur mon parcours.
5
6   ## Remise en question et stage
7 Mon expérience de stage a renforcé cette réflexion et m'a montré que j'avais encore beaucoup à apprendre avant d'entrer pleinement dans le monde professionnel.
8
9   ## Objectifs pour la troisième année
10 Pour la troisième année, je me fixe des objectifs clairs :
11 * **Technique** : Renforcer mes compétences.
12 * **Académique** : Améliorer mes résultats pour viser l'excellence.
13 * **Personnel** : Gagner en confiance et maintenir un niveau régulier.
14
15 Cette période est abordée avec **lucidité** et **détermination**.
```

La partie la plus importante a été le "chunking", c'est-à-dire le découpage du texte. Pour que l'IA soit précise, j'ai structuré mes fichiers avec des titres clairs. De cette façon, le programme peut découper le texte en petits morceaux cohérents. Si quelqu'un pose une question sur mon stage, l'IA sait exactement quel morceau de texte récupérer au lieu de lire tout mon CV d'un coup.

## La partie technique et la recherche hybride

Pour stocker tout ça, j'ai utilisé une base de données vectorielle chez Upstash. C'est là que mes textes sont transformés en vecteurs (des listes de nombres) pour que la machine puisse les comparer.

Item	Score	
ID : 02_Formation.md_1	0.033	
DATA : ## Présentation générale Cette deuxième année de BUT Science des Données m'a permis de consolider les acquis de la première année tout en approfondissant des notions plus...		
VECTOR : [-0.008644104, -0.01209259, -0.03768921, 0.012565613, -0.025039673, -0.035095215, -0...		
SPARSE VECTOR : { "indices": [ 1269207877, -486708, -121218010, -1933587122, 1931152241, 543259558, ...		
METADATA : <a href="#">▶ source</a>		

J'ai choisi une recherche "hybride" : elle combine une recherche par sens (comprendre l'idée générale) et une recherche par mots-clés (pour ne pas rater des termes précis comme "RShiny"). Pour le cerveau de l'IA, j'ai utilisé le modèle Llama 3.3 via Groq. C'est un modèle très rapide qui permet d'avoir une réponse presque instantanée, ce qui rend la discussion beaucoup plus agréable pour l'utilisateur.

## L'interface et les petits plus pour l'utilisateur

Pour l'application web, j'ai utilisé Streamlit. C'est un outil très pratique qui permet de faire une interface propre sans avoir besoin de coder en HTML ou CSS. J'ai essayé de rendre le tout très simple d'utilisation.

The screenshot shows a user interface for an AI assistant named Julien Forestier. On the left, there's a sidebar with a 'Contact' section containing links to LinkedIn, GitHub, and Email. Below that is a button to 'Télécharger mon CV'. The main area features a large title 'Assistant Portfolio - Julien Forestier'. A message from the AI says 'Bonjour ! Je suis l'assistant de Julien. Cliquez sur une suggestion ou posez-moi une question.' A question card asks 'Quels sont tes projets techniques les plus marquants (R, Scala, Python) ?'. The AI's response card states that Julien has successfully completed several technical projects, notably:

- Une régression linéaire multiple sur un jeu de données réel avec RStudio dans le cadre de son cursus en BUT Science des Données.
- Il a également appris le langage Scala dans un environnement Linux et a travaillé sur des projets impliquant la programmation avancée en Python.

J'ai ajouté des boutons de suggestions en haut de la page pour aider les gens qui ne savent pas quoi demander. Par exemple, en un clic, ils peuvent voir mes projets ou mon parcours. J'ai aussi intégré un bouton pour télécharger mon CV en PDF. Ce bouton est "intelligent" : il apparaît dans la barre de côté, mais il peut aussi s'afficher directement dans le chat si on demande à l'IA "je peux voir ton CV ?".

## **Les problèmes et autres contre temps**

Tout n'a pas marché du premier coup. J'ai eu pas mal de soucis avec les versions de Python et des bibliothèques qui ne voulaient pas s'installer correctement. J'ai aussi dû passer du temps à régler les problèmes de clés API qui n'étaient pas reconnues à cause du chemin de mon fichier de configuration. Ces erreurs m'ont appris à être plus rigoureux dans l'organisation de mes dossiers et dans la gestion de mon environnement de travail. Finalement, ce projet m'a permis de voir toute la chaîne de création d'une application IA : de la rédaction des textes à la mise en ligne, en passant par la base de données. C'est une super expérience qui montre concrètement ce qu'on peut faire aujourd'hui avec les LLM.