

SUPINFO Mentor AI

Assistant Académique Intelligent

Bienvenue à l'atelier !
Je suis le Professeur Lapin.

Laissez-moi vous guider à travers
l'architecture de votre nouvel assistant
intelligent, conçu pour naviguer dans
le savoir de SUPINFO.



Basé sur RAG (Retrieval-Augmented Generation)
• Développé par MEVENGUE Franck.

La Mission : Dompter la Connaissance

Le Problème: Les étudiants ont des milliers de questions sur les programmes, les admissions et la pédagogie.

La Solution* : Une IA capable de lire et comprendre plus de **99 documents officiels** instantanément.



Capacités Clés

- Chat intelligent contextuel
- Recherche sémantique avancée
- Personnalités d'assistants configurables

La Vitrine : Une Interface Caméléon

****Design System**** : Une UI moderne et responsive (Radix UI + TailwindCSS).

Thèmes Multiples

L'interface s'adapte à la personnalité.

- ☀️ Light / 🌙 Dark
- 🌸 Pastel / 💐 Girl / ⚪ Boy
- 🤖 Cyber (Mode Hacker)

Le Gardien : Authentification & Rôles

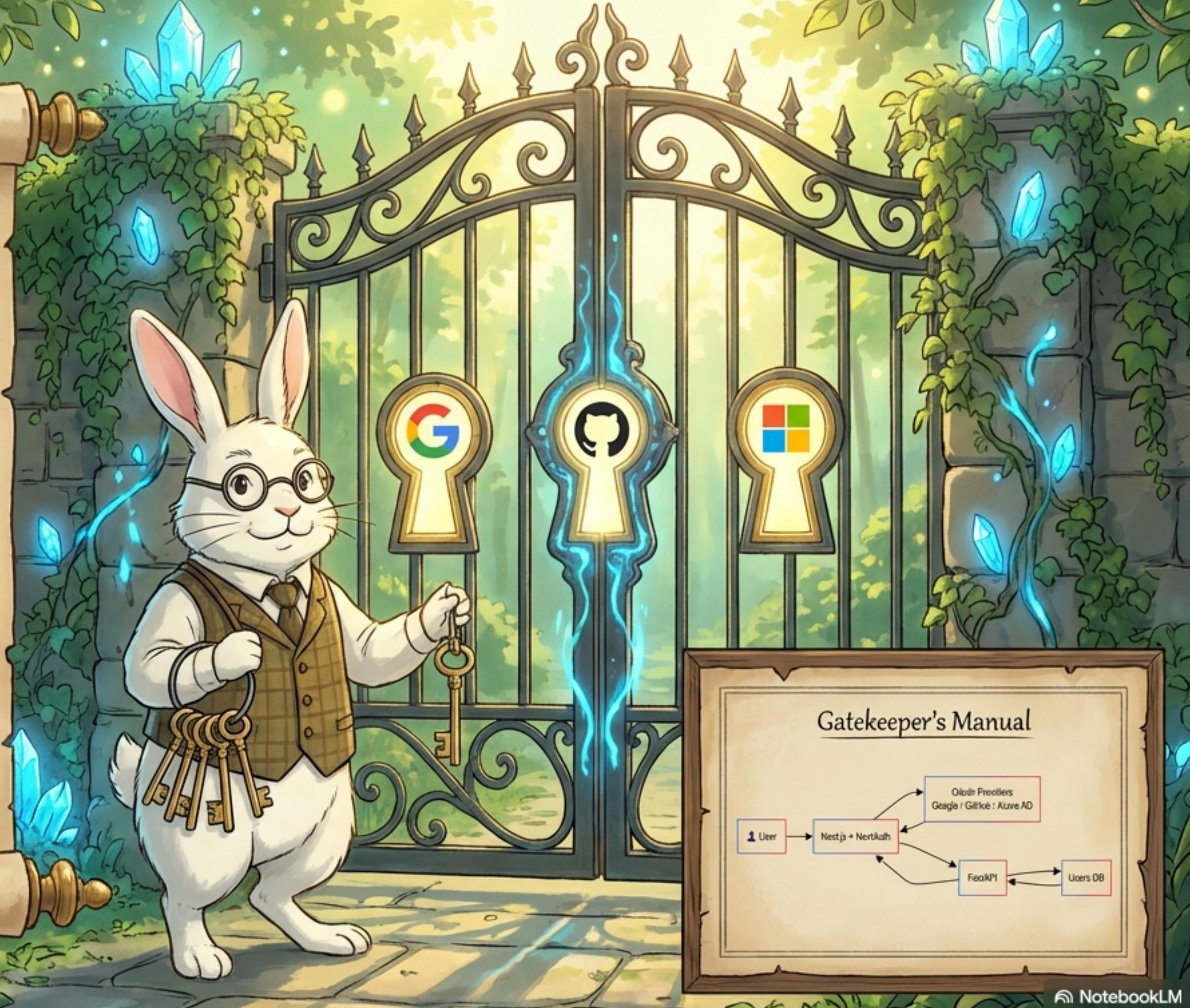
Accès Sécurisé (NextAuth.js) :

🔑 OAuth : Google, GitHub, Microsoft Azure AD.

✉ Credentials : Email & Mot de passe (chiffré avec bcrypt).

Hiérarchie des Utilisateurs :

1. **Guest**: Accès limité (mode découverte).
2. **User**: Historique de conversation, prompts favoris.
3. **Admin**: Dashboard complet et notifications.

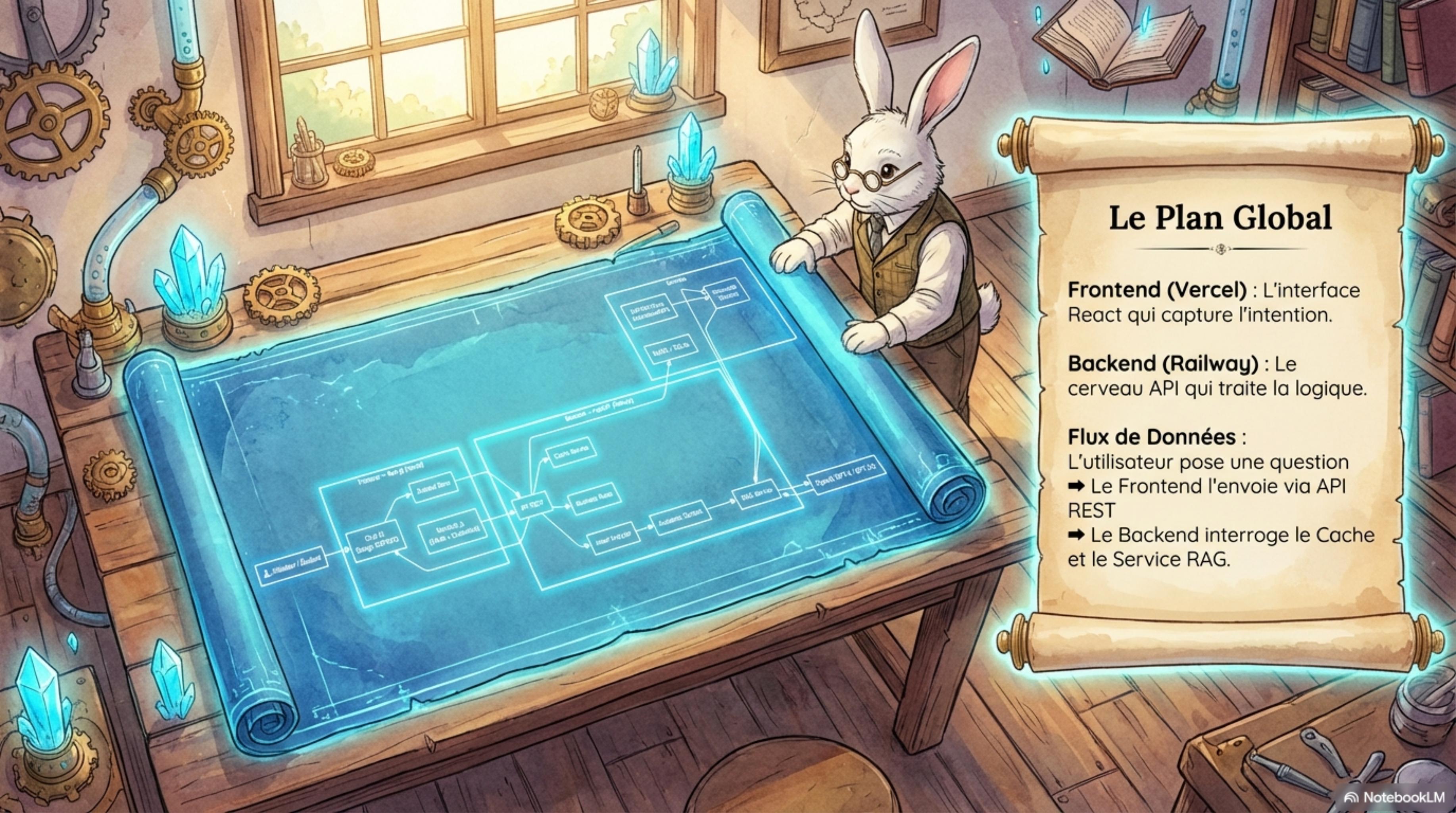


Le Plan Global

Frontend (Vercel) : L'interface React qui capture l'intention.

Backend (Railway) : Le cerveau API qui traite la logique.

Flux de Données :
L'utilisateur pose une question
► Le Frontend l'envoie via API REST
► Le Backend interroge le Cache et le Service RAG.





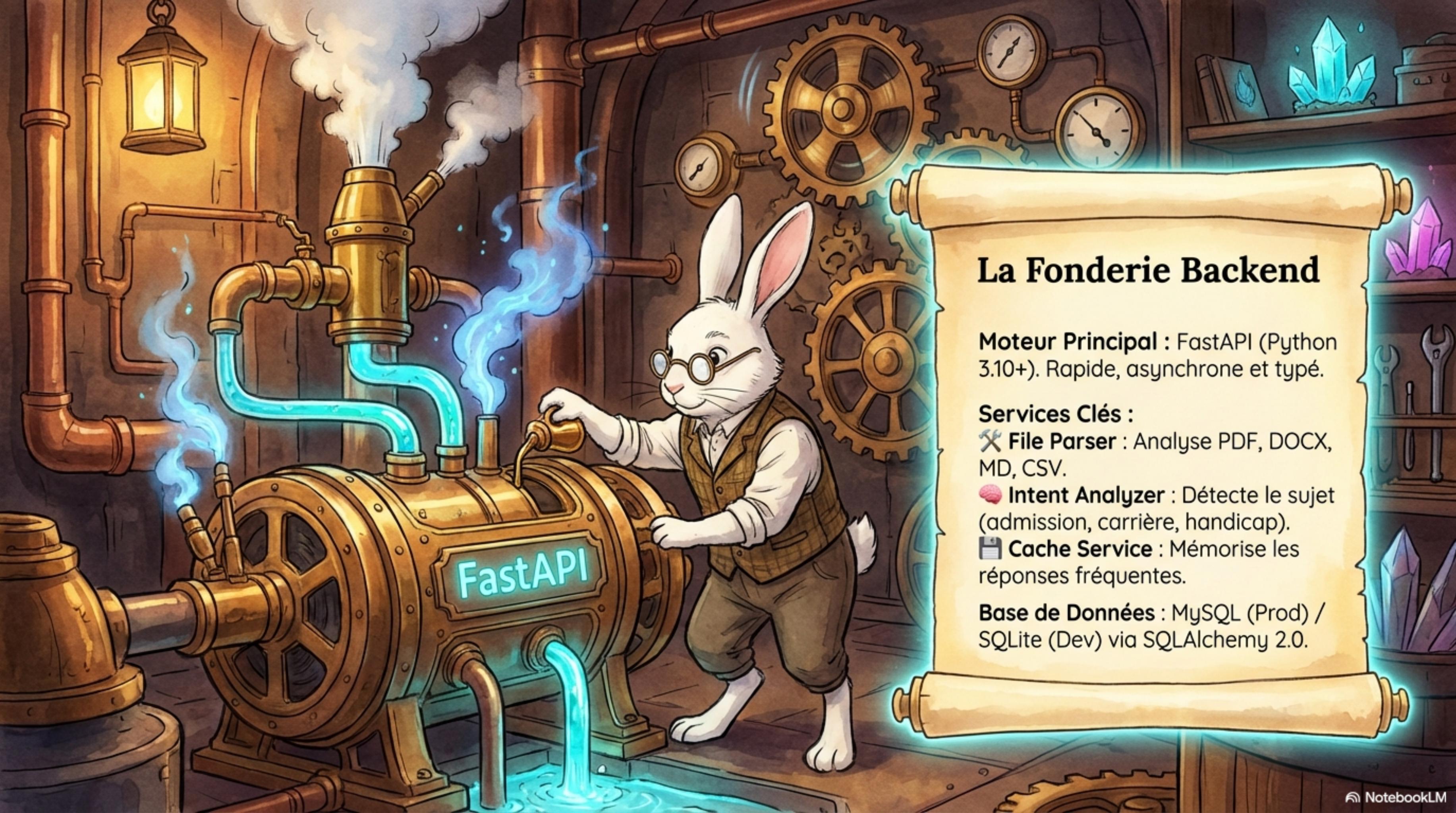
L'Atelier Frontend

****Technologies** :** Next.js 14 (App Router), TypeScript, TailwindCSS.

****Gestion d'État (Zustand)** :**

- 📦 `chat-store` : Gère les messages et le mode de pensée.
- 👑 `avatar-store` : Gère les personnalités de l'IA.
- 📁 `folders-store` : Organise les conversations.

****Composants**:** Architecture modulaire avec Radix UI et Lucide React.



La Fonderie Backend

Moteur Principal : FastAPI (Python 3.10+). Rapide, asynchrone et typé.

Services Clés :

🛠 **File Parser** : Analyse PDF, DOCX, MD, CSV.

🧠 **Intent Analyzer** : Déetecte le sujet (admission, carrière, handicap).

💾 **Cache Service** : Mémorise les réponses fréquentes.

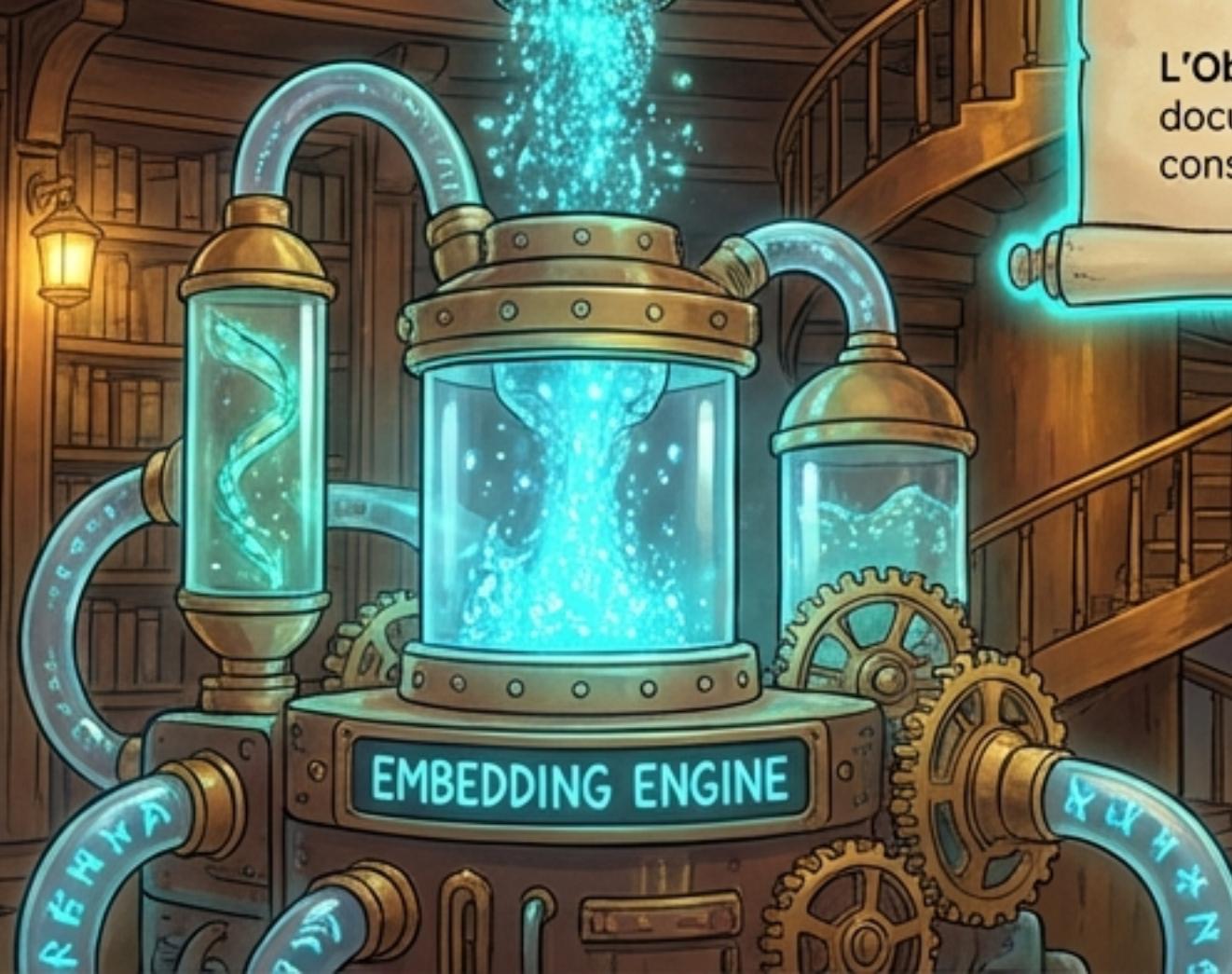
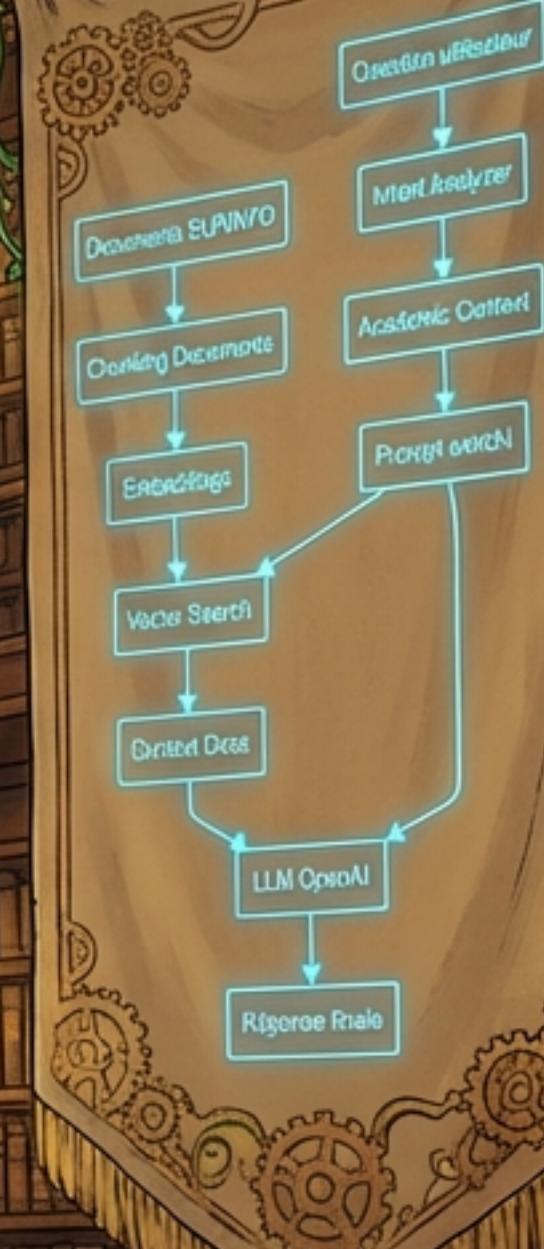
Base de Données : MySQL (Prod) / SQLite (Dev) via SQLAlchemy 2.0.

La Grande Bibliothèque (RAG)

Le Processus d'Ingestion :

1. Documents SUPINFO : Markdown, PDF, Texte.
2. Chunking : Découpage en morceaux de 1000 tokens (overlap 200).
3. Embeddings : Transformation du texte en vecteurs mathématiques (OpenAI/HuggingFace).

L'Objectif : Transformer des documents statiques en une mémoire consultable.





La Mémoire de Cristal (ChromaDB) in Lora

****Recherche Vectorielle** :**

Utilisation de **ChromaDB** pour trouver les passages les plus pertinents sémantiquement.

****Enrichissement :** L'analyseur d'intention ajoute le contexte académique. Le système récupère les 'Chunks' similaires à la question.

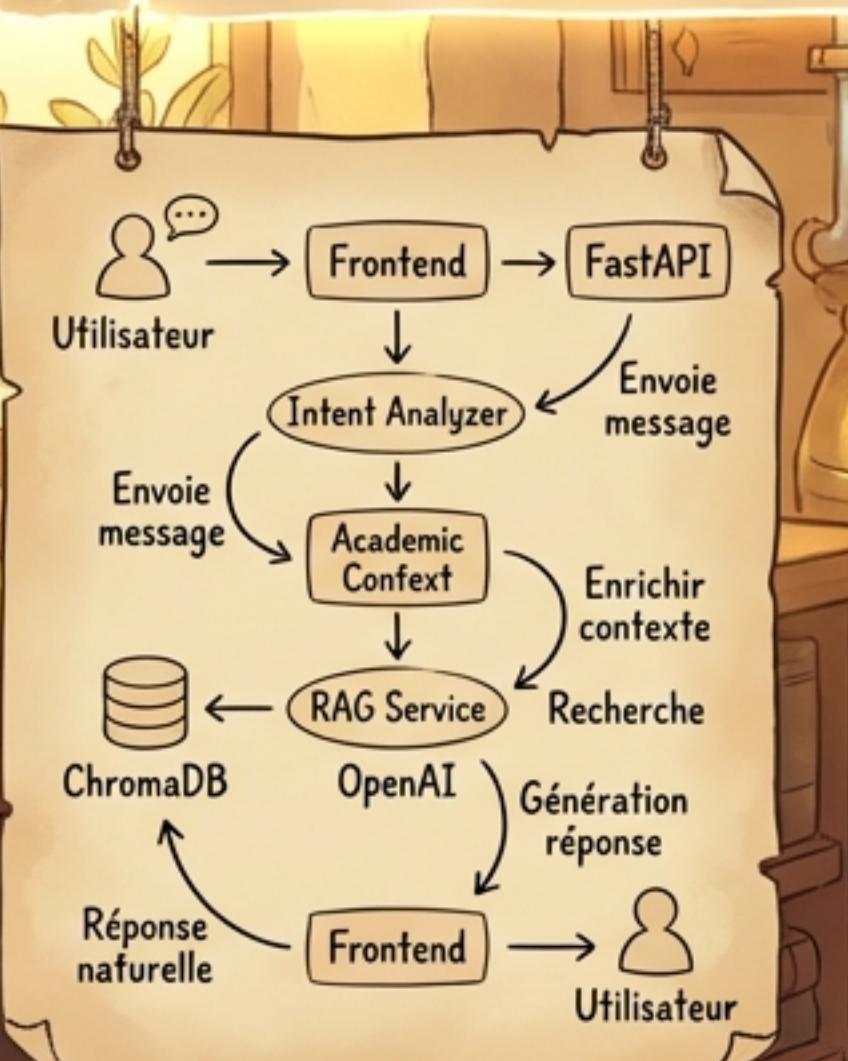
Résultat : Une sélection précise de connaissances sans hallucination.

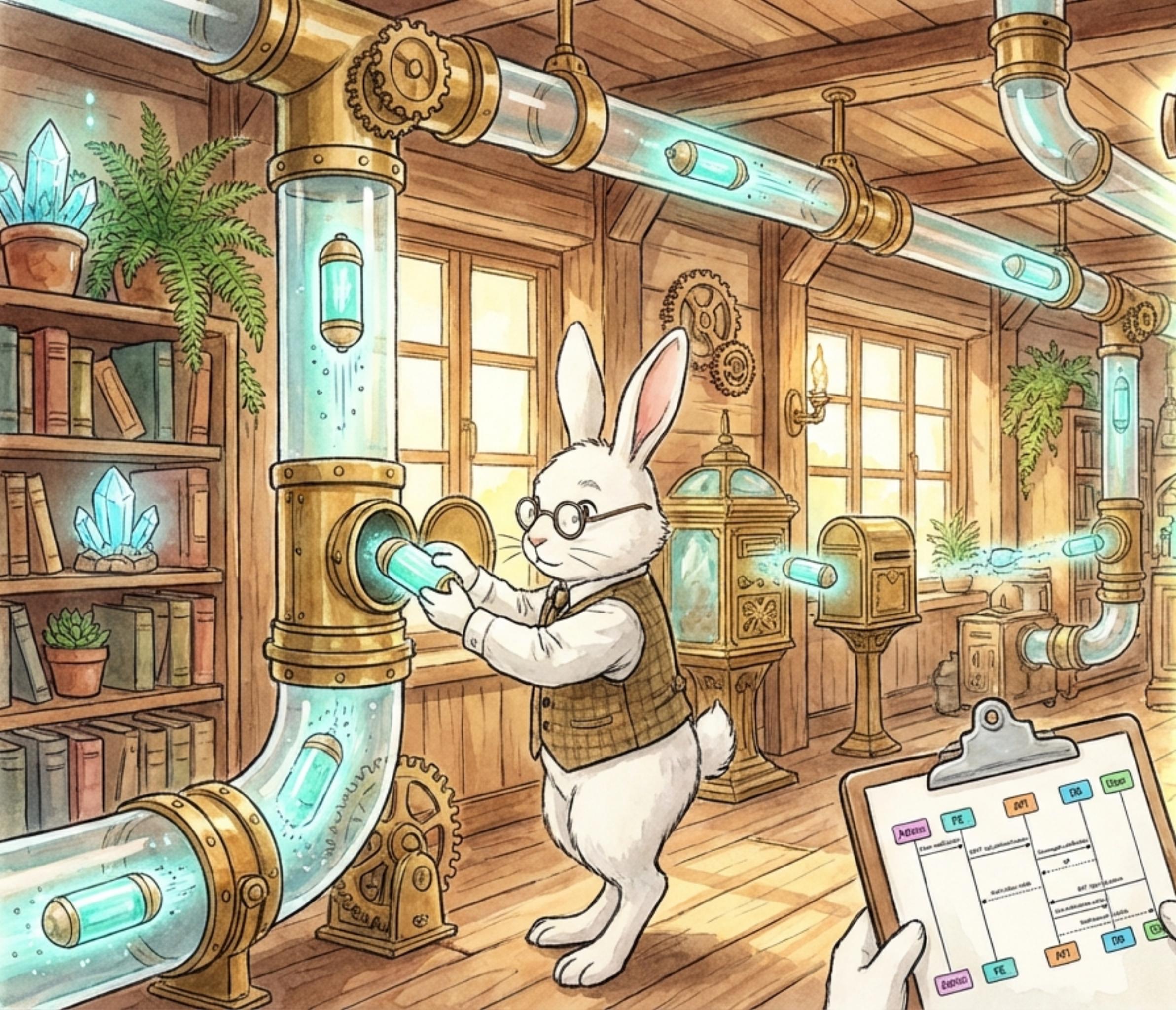
L'Alchimie de la Réponse (LLM)

La Recette : Question Utilisateur + Contexte Académique (Docs) + Prompt Système.

Le Cerveau (OpenAI) : GPT-4 ou GPT-3.5 turbo assemble ces éléments.

Sortie : Une réponse naturelle, sourcée et précise, renvoyée au Frontend via l'API Chat.





Le Système de Communication

Notifications Admin : Les administrateurs peuvent diffuser des messages importants.

Le Flux :

1. Admin crée une notif ('POST /api/admin/notifications').
2. L'API la sauvegarde en base de données.
3. Le Frontend récupère les notifications actives ('GET /api/notifications').

Usage : Annonces de maintenance, nouveautés pédagogiques.

Vercel



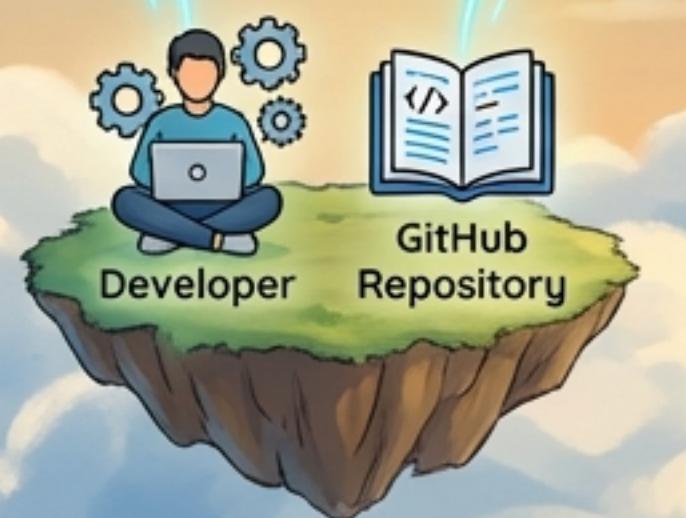
Railway



Déploiement dans les Nuages

Frontend Vercel : Déploiement automatique via GitHub. Optimisation des assets et Edge Network.

Backend Railway : Service Dockerisé (Uvicorn). Hébergement de la base MySQL et ChromaDB. Variables d'environnement sécurisées.



La Tour de Contrôle (Admin)

Dashboard Complet :

- 👤 Gestion des utilisateurs et rôles.
- 📁 Vue globale sur toutes les conversations.
- ⚙️ Configuration des Avatars et Prompts.

Parsing de Fichiers : Upload et traitement centralisé des documents.

Scalabilité : Architecture prête pour l'ajout de nouveaux modèles.



Fin de la Visite

Stack Technique Résumée :

- Next.js 14
- FastAPI
- LangChain
- ChromaDB
- OpenAI

Auteurs : MEVENGUE Franck
(Développeur Principal).

Remerciements : SUPINFO pour les données, et la communauté Open Source.

