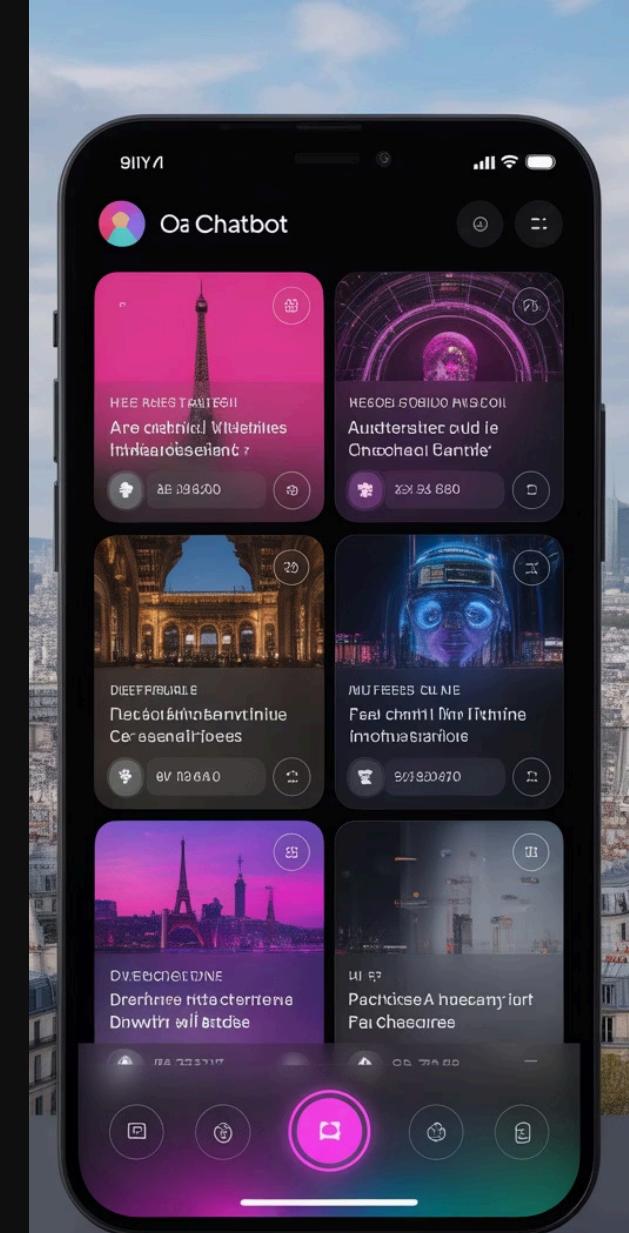


Puls-Events AI : Assistant culturel intelligent

POC d'un assistant IA de recommandation d'événements culturels à Paris (OpenAgenda + Mistral + FAISS)

Abdourahamane LY Data Scientist freelance, spécialisé en NLP & systèmes RAG



Contexte

POC d'un système RAG (Retrieval-Augmented Generation) combinant OpenAgenda, FAISS et Mistral pour répondre intelligemment aux questions sur les événements culturels parisiens.

Contexte

Puls-Events souhaite proposer un assistant IA capable de comprendre et répondre aux questions des utilisateurs en langage naturel

Mission

Concevoir et livrer un POC complet démontrant la faisabilité technique d'un assistant intelligent connecté à des données réelles

Rôle

Data Scientist freelance – conception end-to-end exploitable par les équipes produit et marketing

Objectifs du POC

Objectifs techniques

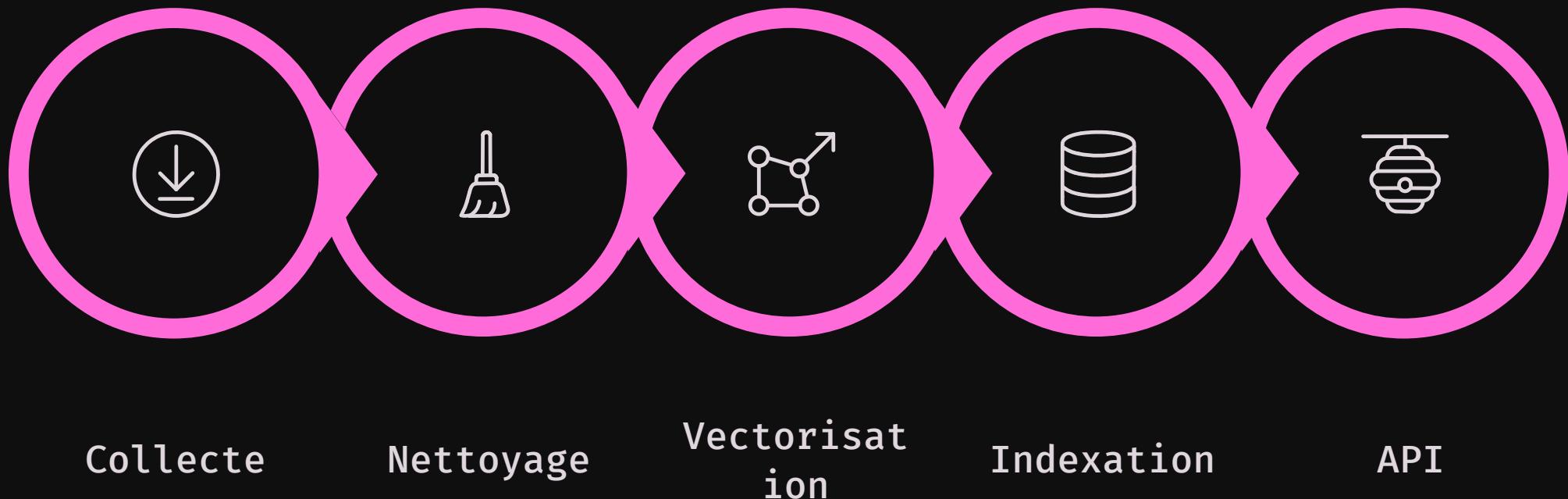
- Construire un pipeline RAG complet et robuste
- Intégrer OpenAgenda, FAISS, LangChain et Mistral
- Exposer le système via une API REST sécurisée
- Permettre la reconstruction automatique des données

Objectifs métier

- Répondre aux questions culturelles en langage naturel
- Produire des réponses fiables, contextualisées et sourcées
- Démontrer la valeur ajoutée pour la plateforme Puls-Events
- Offrir une expérience utilisateur fluide et intuitive

Architecture globale du système

Un pipeline modulaire conçu pour être simple, maintenable et prêt pour l'industrialisation.



Cette architecture permet une évolution progressive du POC vers un système de production tout en conservant la simplicité et la maintenabilité.



Collecte et Préparation des données

1

Source API

Connexion à l'API OpenAgenda avec filtrage géographique (Paris + 20 km) et temporel (1 an passé + 1 an futur)

2

Récupération

Collecte paginée avec gestion des offsets pour récupérer l'ensemble des événements disponibles

3

Nettoyage

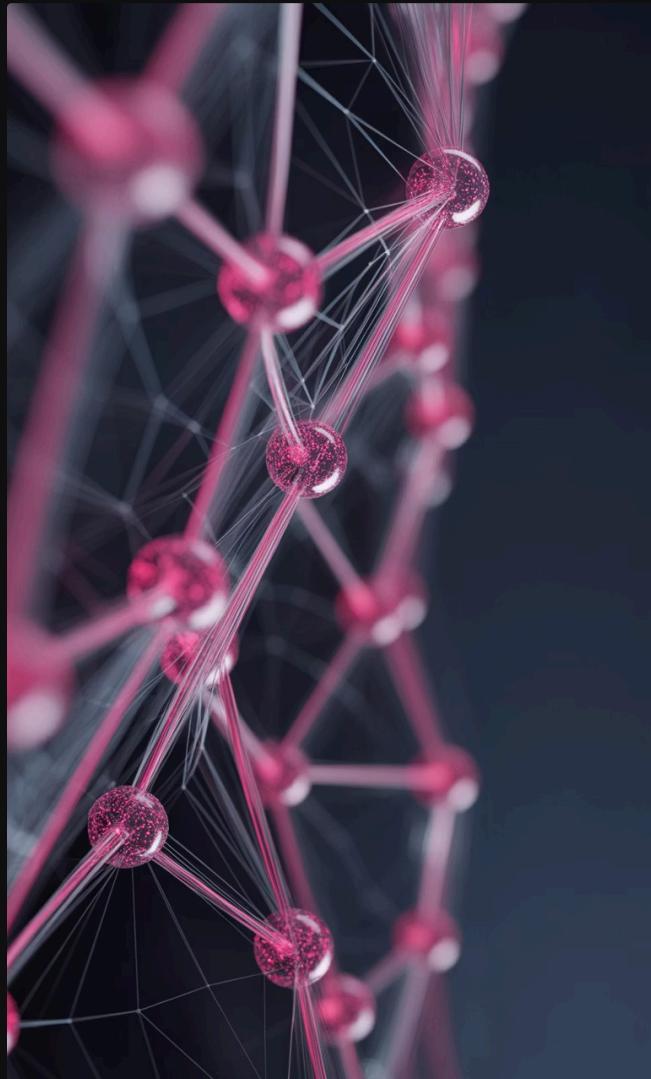
Suppression des doublons, filtrage des descriptions trop courtes, extraction du contenu français uniquement

4

Structuration

Export dans events_raw.csv — dataset propre, structuré et prêt pour l'indexation RAG

Vectorisation et Indexation



01

Transformation

Conversion des événements en documents textuels riches et structurés

02

Découpage

Segmentation intelligente en chunks pour optimiser la recherche sémantique

03

Embeddings

Génération de vecteurs via all-MiniLM-L6-v2, modèle léger et performant

04

Indexation

Stockage dans FAISS (FlatL2) et persistance dans vectorDB

Une base vectorielle rapide, compacte et parfaitement adaptée à la recherche sémantique en temps réel.

Chaîne RAG : LangChain et Mistral



Question utilisateur

L'utilisateur formule une question en langage naturel sur les événements culturels

Recherche vectorielle

FAISS identifie les chunks les plus pertinents via similarité sémantique

Construction du contexte

Les documents pertinents sont assemblés dans un prompt enrichi et structuré

Génération Mistral

Le modèle Mistral (small ou large) génère une réponse naturelle et contextualisée

Réponse avec sources

Retour de la réponse accompagnée des sources pour garantir la transparence



API FastAPI : Exposition du système



POST /chat

Endpoint principal pour les questions utilisateurs – réponse augmentée et sources – sécurisé via X-API-Key



POST /rebuild

Endpoint admin pour reconstruction complète – CSV et FAISS – pipeline automatisé de bout en bout



GET /

Informations de santé et version de l'API pour monitoring et documentation



Logging propre

Traçabilité complète des requêtes et erreurs



Gestion d'erreurs

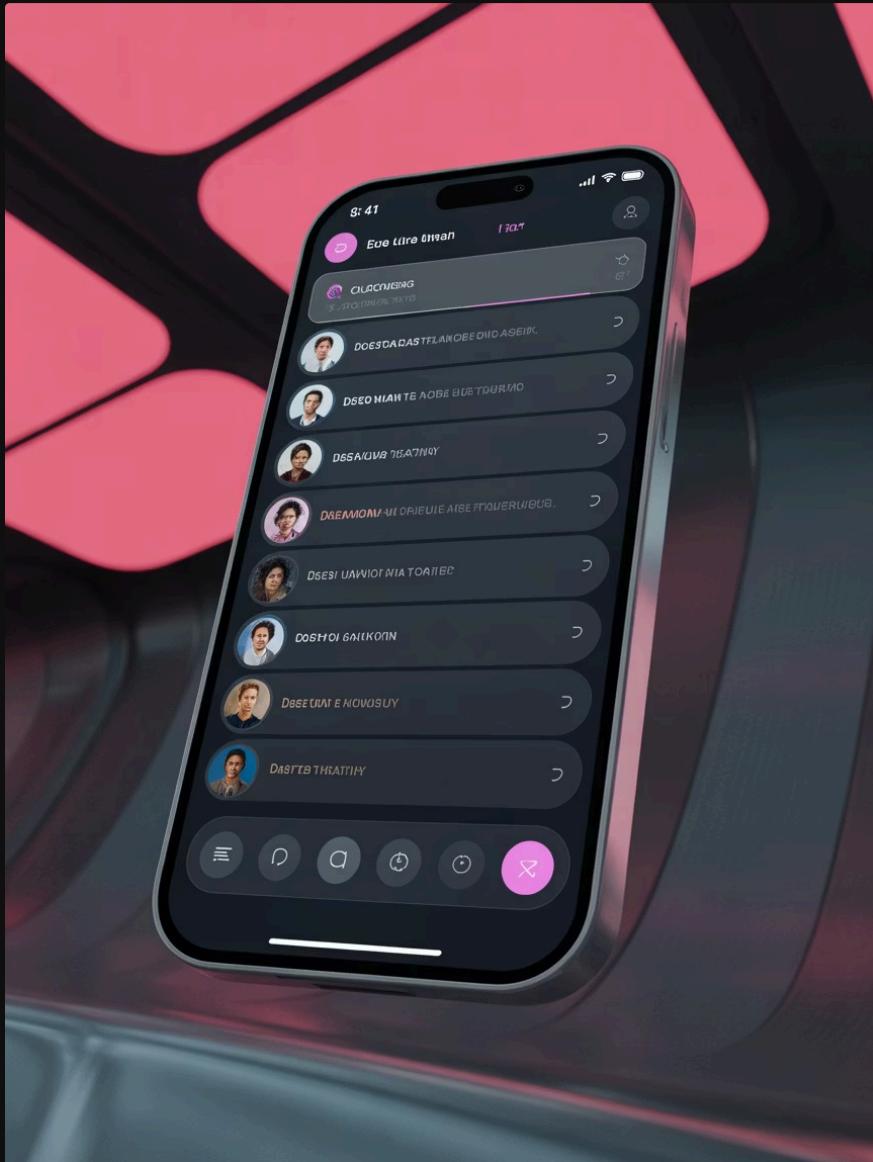
Retours cohérents et informatifs pour tous les cas d'usage



Auto-démarrage

Initialisation automatique du système RAG au lancement

Interface Streamlit



- **Chat moderne avec avatars**

Design épuré et professionnel pour une expérience fluide

- **Sélection du modèle**

Choix entre mode rapide et précis selon les besoins

- **Effet machine à écrire**

Animation naturelle pour une interaction humanisée

- **Sources transparentes**

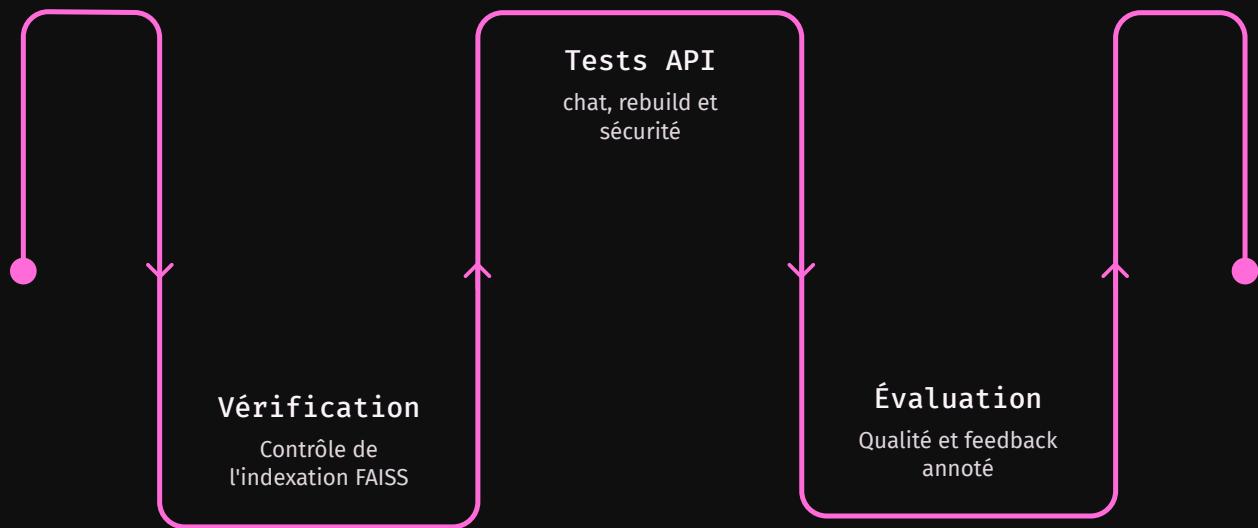
Affichage des références pour chaque réponse générée

- **Feedback utilisateur**

Système de notation / stocké en SQLite pour amélioration continue

Une interface intuitive permettant aux équipes produit et marketing de tester facilement le POC et d'évaluer ses capacités.

Tests, Évaluation et Conteneurisation



Tests techniques

Vérification complète de l'indexation FAISS, tests unitaires des endpoints API, validation de la sécurité et des autorisations

Évaluation qualité

Jeu de test annoté pour mesurer la pertinence des réponses, analyse du feedback utilisateur pour amélioration continue

Docker

Dockerfile optimisé avec uv + pyproject.toml, .dockerignore configuré, API exécutable en un seul docker run

Résultat : Un POC reproductible, portable et prêt pour une démonstration live auprès des équipes.

Résultats et Perspectives

Ce que le POC démontre

- Intégration réussie OpenAgenda → RAG → Mistral
- Réponses fiables, sourcées et compréhensibles
- API exploitable par les équipes Puls-Events
- Interface utilisateur fonctionnelle et intuitive
- Pipeline complet, automatisé et maintenable

Perspectives d'évolution

- Personnalisation par profil utilisateur
- Extension géographique à d'autres villes
- Monitoring et optimisation des performances
- Intégration native dans la plateforme Puls-Events

Conclusion : Le POC valide la faisabilité technique et ouvre la voie à un assistant culturel intelligent pleinement intégré à Puls-Events, capable de transformer l'expérience utilisateur et de valoriser le catalogue d'événements parisiens.