

Jean Savary

Senior Data Scientist & Backend Engineer Python

 jeansvy@gmail.com  linkedin.com/in/jean-savary-806681131

 Madrid, Espagne (Remote + 1 visite/mois en France)

Data Scientist & Backend Engineer avec 6 ans d'expérience en freelance spécialisé en Python, Data Engineering, Machine Learning et architecture backend. Expert en conception d'APIs scalables, pipelines de données, et déploiement cloud (AWS). Passionné par l'optimisation de performances et la création de solutions data-driven robustes.

COMPÉTENCES TECHNIQUES

Backend & Data

Python, Django, FastAPI, Flask, PostgreSQL, Redis, Celery, SQLAlchemy

DevOps & Cloud

AWS (EC2, S3, Lambda, RDS, EKS), Docker, Kubernetes, CI/CD, Terraform, Linux

Data Science & ML

Scikit-learn, Pandas, NumPy, TensorFlow, PyTorch, NLP, Time Series, MLOps

Databases & Tools

PostgreSQL, MongoDB, ArangoDB, Redis, BigQuery, ChromaDB, Airflow

EXPÉRIENCE PROFESSIONNELLE

Co-founder & CTO

Avril 2025 - Présent

HoptyCV • Plateforme SaaS • Remote

- Développement d'une solution complète de génération automatique de dossiers de compétences via IA
- Architecture backend FastAPI avec agents IA pour parsing de CVs, extraction de compétences et scoring
- Pipeline ETL robuste pour traitement et enrichissement de données profils à grande échelle
- Système de templates dynamiques avec génération PDF et adaptation automatique aux offres d'emploi
- Intégration LLMs (OpenAI, Perplexity) pour analyse sémantique et matching de compétences
- Implémentation d'orchestration de tâches asynchrones pour traitement batch de CVs

Stack: Python, FastAPI, PostgreSQL, Redis, Celery, Airflow, AWS (S3, Lambda, EC2), Docker, LLM APIs, Supabase

Founder & CTO

Septembre 2024 - Mars 2025

Reevu • CRM pour Créateurs YouTube • Remote

- Conception et développement du backend pour plateforme d'analytics vidéo YouTube
- Architecture ETL pour ingestion temps-réel de données YouTube (commentaires, métriques) via API
- Modèles ML pour analyse de sentiment et extraction d'insights des commentaires
- Base de données optimisée (PostgreSQL) pour stockage et requêtes rapides sur millions de commentaires
- Pipeline de traitement NLP avec vectorisation et clustering pour segmentation d'audience
- Webhooks et système de notifications pour alertes temps-réel

Stack: Python, FastAPI, PostgreSQL, Redis, OpenAI API, Embeddings, NLP (NLTK, spaCy), AWS (EC2, RDS, S3), Docker, Celery

CTO

Avril 2024 - Décembre 2024

Seeqle • Paris (Remote)

- Refonte complète de l'infrastructure backend et architecture micro-services
- Conception et implémentation d'une base de données de personas optimisée pour recherche de candidats
- Pipeline Airflow pour parsing automatique de CVs, extraction d'entités et scoring de compétences via LLM
- Infrastructure cloud avec Railway et AWS pour déploiement scalable et monitoring
- Encadrement des équipes Backend (2-3 personnes) et Data (2-3 personnes) - architecture et code review
- Optimization des performances: réduction du temps de processing CV de 30s à 2s, latence API de 5s à 500ms
- Contribution à la stratégie produit et définition de la roadmap technique

Stack: Python, FastAPI, PostgreSQL, Redis, Airflow, LLM APIs (OpenAI), Railway, AWS (EC2, RDS, S3), Docker, Kubernetes

Data Scientist & Backend

Mars 2024 - Juin 2024

Animaj • Audiovisuel • Paris (Freelance)

- Détection automatique d'épisodes au sein de vidéos de compilations à grande échelle (~50K vidéos)
- Pipeline de traitement vidéo: téléchargement avec Pytube, extraction frames, encodage vectoriel
- Algorithme d'embedding personnalisé pour reconnaissance de contenu visuel et détection d'épisodes
- Infrastructure de stockage distribué avec AWS S3 et vectorstore ChromaDB pour similarité rapide
- Système de scoring intelligent pour classification et regroupement automatique d'épisodes
- Déploiement et orchestration sur AWS EC2 pour processing distribué

Stack: Python, OpenCV, Embeddings, ChromaDB, AWS (S3, EC2), Pytube, NumPy, Pandas, Scikit-learn

Team Leader | Data Science & Backend Python

Décembre 2022 - Septembre 2024

Diagoriente • Paris

- Management d'une équipe de 4 data scientists/engineers sur des projets backend Python et ML
- Développement d'une architecture microservices avec FastAPI et Django
- Création et maintenance d'une base de données graph (ArangoDB) pour compétences professionnelles
- Conception d'APIs REST pour moteur de recommandation basé sur ML

Data Scientist

Septembre 2018 - Mars 2020

Evaneos • Région de Paris

- Outil de taggage automatique NLP des données textuelles relatives aux projets de voyage
- Analyses ad-hoc pour vente, marketing, acquisition, produit
- POC sous Flask et développement de plateforme de reporting interne
- Scoring sessions online et modèle de pricing (data mining, machine learning)
- Analyses de clustering (clients, destinations) et causalité
- Formations de l'équipe Data: SQL, BigQuery et Python

Stack: Python, Flask, BigQuery, Jupyter, Scikit-learn, NLP (TF-IDF, word2vec), Dask, Google Analytics

Autres expériences:

Backend Engineer Python	MARBL	Juillet 2021 - Octobre 2022
Lead Teacher Data Science	Le Wagon	Juin 2021 - Décembre 2021
Data Engineer	Stealth	Janvier 2021 - Mars 2022
Data Scientist & Backend Engineer	Innov-Training	Juillet 2020 - Décembre 2020
CTO	Bouge	Juin 2020 - Mars 2021
Data Analyst Operations	MoovOne	Septembre 2021 - Novembre 2021

FORMATION

Diplôme d'Ingénieur - Computer Science

IMT Atlantique • 2017 - 2020

Master of Science - Computer Science

École Polytechnique de Montréal • 2018

LANGUES

Français (Natif) Anglais (Professionnel) Espagnol (Professionnel)