

_ CURRICULUM

Baptiste Marc Mokas

SENIOR DATA SCIENTIST MLOPS | ENTREPRENEUR | RESEARCH SCIENTIST

Languages : English / French / Greek

born 03/28/1994 (31 yo)

250, rue de Carnoy
59130, Lambersart

baptiste@weeki.io ✉
+33 7 69 08 54 19 ☎

FRANCE

weeki.io/baptiste-mokas/ 🌐

linkedin.com/in/baptiste-mokas/ 📄

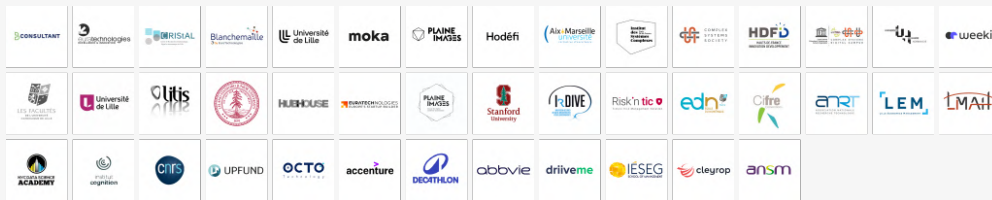


Bio

Versatile professional scientist with over 8 years of experience as a MLops Data Scientist and as a dedicated Researcher in Artificial Intelligence, my goal is to help companies structure and leverage their data for informed strategic decisions.

In addition, I serve as a Professor of Data Science at the Catholic University of Lille, where I mentor and guide the next generation of data scientists.

I am also the founder of Weeki, an AI-powered SaaS platform, which boasts thousands of teams gain productivity. The platform enables them to centralize, structure, and visualize their knowledge within an interactive collaborative virtual whiteboard. Weeki streamlines the organization of documents, data, and tools and goes further by embedding pipeline & AI agent execution in the knowledge workspace.



Professional Positions

Lead Senior Data Scientist & MLops at Decathlon Technology
Teacher in Data Science & Research associate at Lille Catholic University
Founder of Weeki, an AI-powered SaaS Platform for Knowledge Management

Education

PhD student in Mathematics and Computer Science — Le Havre University (2019 - 2021)
Master of Mathematics and Computer Science — Lille University (2017 - 2019)
Diplomas in Entrepreneurship and Data science — Stanford & New York (2021 - 2023)

Leading Skills

AI / Machine Learning Mathematical Modeling Pedagogical Engineering
Scientific Research Data Analysis Cloud / Data Engineering UX/UI Design
Software Development Workflow & Team Management Tech Startups

Key Success Factors / Publication & Awards

500k€ fund raised as a solofounder / +1K users / 5 years Incubated in Euratechnologies
My music studio was featured in MusicTech magazine
Entrepreneurship Prize of Lille University

Freelance services:

Data processing

- Python (Poetry) / R ██████████
- ETL (Airflow, DBT, Fivetran) . . . ██████████
- Data Analysis / EDA ██████████
- Web Scraping (Python) ██████████
- Stream (Beam, Dataflow, Kafka) ██████████
- Cassandra, Hadoop / Spark ██████████
- PySpark ██████████

GenAI & LLMs

- LLMs / ChatGPT API ██████████
- Prompt Eng (CoT Calling) ██████████
- RAG (LangChain, LlamaIdx) ██████████
- Vector DB (FAISS, Chroma) ██████████
- Agents Reasoning & MCP ██████████
- Guardrails & Monitoring ██████████

Prédiction & Classification

- Mathematical Modeling ██████████
- ML & Statistical Models (50+) ██████████
- Deep Learning / Tensorflow ██████████
- Pytorch, Scikit-learn, XGBoost ██████████
- Computer Vision / OCR ██████████
- Time Series / Forecast ██████████
- NLP / Data-to-Texte ██████████

Data Visualization

- Dashboards (Tableau, PowerBI) ██████████
- Plots / Graphics (25+) ██████████

Product development

- UX/UI Prototype (XD, Figma) ██████████
- Frontend (React, JSX, SCSS) ██████████
- Backend APIs (Flask, FastAPI) ██████████
- SQL / NoSQL (MongoDB) ██████████
- Nocode (Make, Zapier, N8n) ██████████

Production / ML-ops

- CI/CD (Jenkins, Git, Terraform) ██████████
- Cloud (AWS / GCP / Azure) ██████████
- AWS (EC2, S3, RDS, Lambda) ██████████
- ML pipeline (Ray, Kubeflow) ██████████
- Databricks & MLflow ██████████
- Dataprep, Dataiku & Alteryx ██████████
- Docker, Kubernetes & Ansible ██████████
- Cybersecurity (OWASP, ISO) ██████████
- ELK Stack (Elastic, Kibana) ██████████

Project management

- Agile / Scrum management ██████████
- Literature Review ██████████
- Experimental Design ██████████

PORTFOLIO

SASU WEEKI | Baptiste Marc MOKAS

+33 (0)7 69 08 54 19

[Schedule a meeting with me](#)

baptiste@weeki.io

TABLE DES MATIÈRES

ADEO LEROY MERLIN – MLOPS GENAI & LLM PRODUCTION	2
PACKITOO – OPTIMISATION DE PRICING (DEVIS)	3
ABBVIE PHARMACEUTIQUE – AI MARKETING ASSETS CONTENT AUTOMATION	4
DRIIVEME TECHNOLOGY – PRICING OPTIMIZATION	5
ANSM (AGENCE NATIONALE DE SÉCURITÉ DU MÉDICAMENT) GENAI DOCUMENTAIRE & SUPERVISION INTELLIGENTE	6
QUALINTRA DATA SCIENCE RH – ENGAGEMENT SCORING & MODEL EXPLAINABILITY	7
WEEKI.IO TECHNOLOGY – AI AGENT & MLOPS	8
DECATHLON OCTO TECHNOLOGY – SUPPLY CHAIN FORECAST	9
UNIVERSITÉ CATHOLIQUE ENSEIGNANT DATA SCIENCE & IA	12
WEEKI.IO SAAS – KNOWLEDGE MAPPING	13
NOE ML POUR LA REVUE DE LITTÉRATURE CLINIQUE	14
PHD STUDIES RESEARCH & DATA SCIENCE	15
SCALAB / CRISTAL LAB RESEARCH – MACHINE LEARNING SOCIAL BEHAVIOR	16

Lead MLOps Senior – Exploitation & mise en production ML/GenAI

Mission hybride, France. Anglais.

Mission

- Mise en production ML & GenAI : industrialiser les POC/MVP des Data Scientists jusqu'à la prod sur GCP.
- Observabilité & fiabilité : logs, métriques, traces, SLI/SLO/SLA avec Datadog ; alerting & runbooks.
- Sécurité & DevSecOps : gestion des secrets (Vault), durcissement des images, politiques de conformité.
- Orchestration & CI/CD : Kubernetes (containerisation, autoscaling), workflows GitHub Actions, artefacts & stacks (ZenML).
- Performance & FinOps : optimisation latence/disponibilité, rationalisation des coûts d'inférence et d'infra.
- Documentation & exploitation : doc technique rigoureuse (HSV/TSD), playbooks d'exploitation, bonnes pratiques de code.
- Qualité modèle : suivi dérive data/modèle, KPI de performance, tests unitaires & d'intégration.

Contexte & Environnement

- Secteur : Retail & e-commerce.
- Contexte business : mise en production de multiples produits alimentés par LLM pour suivre leurs performances et métriques.
- Cas d'usage clés : (1) génération automatisée de descriptions produits ; (2) chatbot basé sur RAG ; (3) pipeline de scraping pour cartographier signaux/indicateurs financiers via APIs de recherche Google ; (4) génération automatique de matrices SWOT à partir de données financières ; + autres POC/MVP.
- Stack & contraintes : Kubernetes sur Google Cloud ; Vault pour les secrets ; Datadog pour logs/metrics/traces ; fortes exigences sécurité/observabilité et de bonnes pratiques (SLI/SLO/SLA) ; documentation soutenue ; accès via VPN & SSO ; anglais de travail.
- Parties prenantes : Métiers, IT, Data, Ops ; l'équipe MLOps/Machine Learning Engineers en interface transverse.
- Équipe projet : squad MLE/MLOps en mode agile, rituels et collaboration quotidienne avec les équipes produit & data.

Objectifs Clés

- Industrialisation à l'échelle : 5 modèles mis en production prioritaires ; 5 modèles additionnels évalués/assainis et prêts pour l'industrialisation.
- Observabilité standardisée : référentiel global ready (acceptance) couvrant logs, métriques, qualité des données, SLI/SLO/SLA.
- Plateforme robuste : pipelines CI/CD de bout en bout, déploiements reproductibles/rollback, monitoring de santé sur K8s.
- Performance & coûts : latence et disponibilité optimisées ; réduction des coûts d'inférence et d'infrastructure (FinOps).
- Sécurité : secrets centralisés, conformité aux politiques internes, auditabilité bout-en-bout.

Technologies & Infrastructure

- Langages & Frameworks : Python ; PyTorch (deep learning) ; RAG ; prompt engineering (Chain-of-Thought) ; tests unitaires & d'intégration.
- LLM & services : Vertex AI (Gemini), OpenAI ; comparaison et évaluation multi-LLM.
- Cloud & Orchestration : Google Cloud Platform ; Kubernetes ; GitHub Actions (workflows CI/CD) ; Docker.
- Données & Pipelines : DBT ; Kafka ; gestion de bases de données selon projet ; ZenML (gestion des stacks & artefacts, observabilité modèles).
- Monitoring & Observabilité : Datadog (logs, métriques, traces) ; tableaux de suivi de performance & dérive.
- Sécurité & Accès : Vault (gestion des secrets), VPN, SSO.

PACKITOO – OPTIMISATION DE PRICING (DEVIS)

Lead Senior Data Scientist – Pricing & Revenue Optimization

Mission — localisation & modalités à confirmer.

Mission

- Optimisation de prix : modéliser la probabilité d'acceptation d'un devis (accepté/non accepté) et recommander le prix optimal maximisant la marge.
- Classification & scoring : XGBoost (logistic) pour estimer la propension d'acceptation par devis.
- Simulations what-if : moteur de scénarios faisant varier le prix de -99 % à +100 % pour explorer le couple taux d'acceptation ↔ marge et tracer les points d'optimum par segment.
- Pipelines data & qualité : nettoyage, transformations, normalisation/standardisation, détection et suppression des fuites (leakage), séparation temporelle stricte.
- Validation robuste : validation croisée temporelle (Time Series CV), suivi train/test et garde-fous de généralisation.
- Auto-ML raisonné : Optuna / Hyperopt pour l'hyperparamétrisation (recherche bayésienne/Tree-Parzen).
- Interprétabilité & confiance : SHAP, LIME, importances locales & globales, rapports d'explication pour les métiers.
- Suivi de la performance : dashboards de métriques (comparaison RMSE train/test, LogLoss, AUC/PR), suivi dérive données/modèle.
- Industrialisation : packaging du modèle, mise en production (batch/API), documentation et playbooks d'usage.

Contexte & Environnement

- Périmètre métier : analyse des devis avec de multiples critères explicatifs (produit, client, contexte commercial, etc.) pour arbitrer prix vs. probabilité d'acceptation.
- Enjeux : fournir aux équipes ventes/finance un simulateur de prix et des recommandations actionnables (couloirs de prix, points d'inflexion) afin d'augmenter la marge sans dégrader la conversion.
- Contraintes : rigueur méthodologique (pas de fuite d'information), temps comme variable structurante (TS-CV), explicabilité forte pour adoption métier.
- Parties prenantes : Sales / Revenu / Finance, Data, IT ; rôle transverse du Data Scientist auprès des décideurs.

Objectifs Clés

- Modèle de propension : score d'acceptation fiable et calibré pour chaque devis.
- Simulateur de pricing : what-if de -99 % à +100 % avec marges projetées et corridors de prix par segment.
- Gains opérationnels : recommandations prêtes à l'emploi (prix cible, intervalle de confiance, trade-offs).
- Robustesse & gouvernance : TS-CV, métriques de généralisation, reportings standardisés et explicables (SHAP/LIME).
- Production & adoption : pipeline de bout en bout (data→modèle→API/batch→dashboard), documentation et formation des utilisateurs.

Technologies & Infrastructure

- Langages & frameworks : Python, pandas, NumPy, scikit-learn (pipelines), XGBoost.
- Optimisation : Optuna, Hyperopt (TPE, early-stopping).
- Validation : Time Series Cross-Validation (type `TimeSeriesSplit`), suivi train/test (RMSE, LogLoss, AUC/PR).
- Interprétabilité : SHAP, LIME (importances locales & globales).
- Pipelines & data prep : feature engineering, normalisation/standardisation, contrôle qualité et traçabilité.
- Mise en production : scoring batch et/ou API, dashboards de performance et dérive (outil interne), documentation.

Lead Senior Data Scientist / AI Solutions Architect – RAG, LLM & Agentic Content Generation

Mission projet (remote/Paris)

Mission

- Automatisation avancée de la génération de documents réglementaires (AMM), supports marketing et présentations PowerPoint à l'aide d'un pipeline agentique RAG/LLM orchestrant tout le cycle de production documentaire, de l'ingestion multi-source à la génération finale contextualisée.
- Développement from scratch d'une application modulaire sur Palantir Code Workspace : ingestion, chunking, embeddings (OpenAI/HuggingFace), vector search (PostgreSQL, SQLAlchemy, Alembic) et gestion fine de la mémoire contextuelle pour des agents réellement autonomes dans leurs actions.
- Intégration de Langchain & LangGraph pour la création d'un véritable collectif d'agents : enchaînement dynamique de tâches, raisonnement multi-étapes (chain of thought), prise d'initiatives sur des scénarios complexes (validation réglementaire, structuration de supports, réponses ad hoc) avec suivi proactif (monitoring Langfuse).
- Automatisation : Création de slides PowerPoint via un moteur agentique de templating : agents capables de générer, organiser et adapter dynamiquement les contenus en blocs/fonctions réutilisables selon les contraintes métier ou conformité réglementaire.
- Prompt engineering agentique avancé : tuning structurel automatisé (température, tokens, paramètres dynamiques) et gestion intelligente des instructions pour guider le comportement des agents selon le contexte documentaire et les exigences du workflow.
- Tests d'orchestration end-to-end : Pytest, validation multi-agents de bout en bout pour garantir robustesse, auditabilité et fiabilité en production.
- Suivi de la performance : monitoring continu par des métriques avancées (precision@k, recall@k, ROUGE, BLEU), dashboards Palantir pour visibilité temps réel, explicabilité des décisions via SHAP, et adaptation continue des agents selon les résultats observés.
- Documentation exhaustive et transfert d'expertise auprès des équipes métier (Marketing/IT) pour maximiser l'adoption et l'autonomie sur la solution.

Contexte & Objectifs

- Pharmaceutique, marketing et affaires réglementaires : orchestrer, fiabiliser et accélérer la création de livrables réglementaires & market assets par des architectures agentiques.
- Objectifs : réduire drastiquement le time-to-market, garantir la compliance, l'auditabilité, la traçabilité, et mesurer la performance de bout en bout grâce à l'agentique AI.

Tech Stack & Méthodologie

- AI / Agentic Intelligence : OpenAI, HuggingFace Transformers, Langchain, LangGraph (workflow multi-agent, orchestration)
- Base & Recherche contextuelle : PostgreSQL, SQLAlchemy, Alembic, vector embeddings
- Programmation : Python objet, design patterns, modularisation et réutilisabilité
- Prompt & Agent engineering : tuning structuré, chain of thought, gestion dynamique des instructions
- Supervision & Évaluation : Langfuse, SHAP, dashboards Palantir, KPIs de retrieval/génération
- Testing & CI/CD : Pytest, tests intégrés, versionning Git
- Visualisation : génération automatique PowerPoint, templating dynamique agentisé

Valeur ajoutée

Déploiement d'une IA agentique orchestrant, contextualisant, corrigeant et adaptant toute la chaîne documentaire via une intelligence autonome, pour un gain opérationnel, qualité et compliance maximal.

DRIIVEME | TECHNOLOGY – PRICING OPTIMIZATION

Lead Senior Data Scientist – Data Valorisation & Machine Learning

Mission courte (5 semaines) / Paris, France (hybride)

Mission

- Audit et cadrage des besoins métier, exploration avancée des bases de données (Databricks, SQL).
- Nettoyage, gestion des valeurs manquantes/outliers, industrialisation des scripts de preprocessing.
- Feature engineering adapté au métier, préparation du socle analytique.
- Mise en œuvre de modèles prédictifs (XGBoost), optimisation et suivi avec MLflow (tracking, versionning).
- Explicabilité des modèles via SHAP pour la transparence métier.
- Simulations what-if et génération de dashboards interactifs pour faciliter la prise de décision.

Contexte & Objectifs

- Mobilité/logistique, données multi-sources et évolutives nécessitant des analyses rapides.
- Objectifs : Fiabiliser la donnée, accélérer l'exploitation analytique, produire des insights opérationnels et faciliter la décision.

Technologies & Méthodologie

- Python (Pandas, scikit-learn, XGBoost, SHAP), SQL
- Databricks, MLflow, scripts versionnés (Git), export Parquet/CSV
- Visualisation : Streamlit, PowerBI/Tableau
- Best practices : tests unitaires, documentation claire, restitution orientée métier

Lead Senior Data Scientist / AI Architect – GenAI, Agentic AI, Surveillance & Automatisation

Mission via partenariat Theodo x Cleyrop (Paris / hybride)

Mission

- Accompagnement des équipes de surveillance réglementaire par l'intégration d'agents intelligents et GenAI pour automatiser et fiabiliser la comparaison des Résumés des Caractéristiques du Produit (RCP) sur base ATC.
- Industrialisation de la veille pharmaceutique : extraction, structuration et analyse automatisée de grands ensembles documentaires réglementaires.
- Développement d'agents d'IA agentic (au-delà des LLMs) orchestrant des tâches complexes : interrogation contextualisée via Retrieval-Augmented Generation (RAG), synthèse automatisée des écarts entre molécules, détection proactive d'anomalies, notifications dynamiques et prises d'initiatives autonomes pour la supervision.
- Vectorisation intelligente et structuration avancée des connaissances par embeddings sur corpus réglementaire, stockage optimisé sur bases vectorielles pour enabling agentic reasoning.
- Mise en œuvre d'algorithmes analytiques d'aide à la décision, détection d'anomalies et scoring d'écarts pour renforcer la fiabilité des analyses documentaires critiques.
- Déploiement de dashboards interactifs pour la restitution dynamique des anomalies et recommandations IA, production automatisée de rapports pour l'aide à la décision et l'orientation opérationnelle.
- Automatisation avancée via architecture objet : génération modulaire de prompts, chaînes de raisonnement complexes (chain-of-thought), tuning de l'orchestration agentic pour maximiser la pertinence et la sécurité opérationnelle.
- Tests robustes (Pytest), monitoring continu des agents (metrics retrieval/generation, SHAP, Langfuse) pour assurer performance, explicabilité et transparence.

Contexte & Objectifs

- Modernisation technologique de l'ANSM en pharmacovigilance et régulation documentaire, pilotée en consortium (Theodo x Cleyrop).
- Objectifs : accélérer la comparaison des RCP, automatiser la surveillance documentaire, réduire le risque d'erreur humaine et augmenter drastiquement l'efficacité et l'autonomie des équipes de supervision.

Technologies & Méthodologie

- AI & Agentic Intelligence : OpenAI, Huggingface Transformers, Langchain, LangGraph (orchestration multi-agent), RAG sur corpus réglementaire
- Stockage & Monitoring : Faiss/PGVector (vector DB), embeddings avancés, PostgreSQL, Langfuse pour le suivi agentic
- Programmation & Orchestration : Python objet, architectures modulaires et réutilisables, SQLAlchemy, Alembic
- Prompt/Agent Engineering : chain of thought, hyperparameter tuning, design multi-fonctions
- Dataviz : Streamlit, PowerBI, générations automatisées de rapports
- Testing & CI/CD : Pytest, orchestration Git
- Métriques & Explicabilité : SHAP, KPIs de retrieval/generation, scoring d'écarts, monitoring explicite agentic

Valeur Ajoutée

Démarche agentic (IA orchestrant, monitorant, décidant et contextualisant dynamiquement les tâches critiques), robustesse analytique, automatisation modulaire et restitution transparente centrée décision métier.

Lead Senior Data Scientist – Modélisation Machine Learning, Explicabilité & Recommandations RH
Mission Data Science / Paris (remote/hybride)

Mission

- **Développement d'un pipeline** : machine learning pour expliquer et optimiser le score d'engagement RH (200 questions, 25 ans de données multi-clients).
- **Migration du modèle statistique (GLM)** : à des algorithmes avancés (XGBoost, LightGBM, Random Forest).
- **Nettoyage et imputation avancée des données** : (66% de valeurs manquantes), détection d'outliers, gestion de la multicolinéarité (corrélation, PCA, régularisation).
- **Explicabilité globale et locale** : SHAP, LIME, analyse de sensibilité (Morris, Sobol).
- **Optimisation** : tuning hyperparamètres (Optuna/Hyperopt), sélection de variables métier, validation croisée, métriques personnalisées (RMSE, MAE, R²).
- **Production de dashboards et rapports** : didactiques avec recommandations RH personnalisées.

Contexte & Objectifs

- **Analytics** : RH, scoring d'engagement collaborateur, explication et hiérarchisation des leviers RH opérationnels.
- **Workflows** : ML robustes et interprétables, restitution claire et actionnable.

Technologies & Méthodologie

Machine Learning : XGBoost, LightGBM, Random Forest

Explicabilité : SHAP, LIME, analyses de sensibilité

Traitement : PySpark dat (Pandas, scikit-learn), pipelines modulaires, imputation avancée

Visualisation : PowerBI/Tableau, dashboards explicatifs

Best practices : documentation exhaustive, restitution orientée client

Senior Lead Data Scientist – Knowledge Management & RAG

Expérience de fondateur / Paris, France.

- **Développement d'agents IA** : intégrant des embeddings vectoriels et des auto-encodeurs pour structurer la carte de connaissances multi-niveaux.
- **Intégration LLM + RAG** : réponses contextualisées (+ attention weight basés sur un transformer)
- **Automatisation des flux** : ingestion, traitement et enrichissement via l'automatisation DAGs.
- **Prompt engineering avancé** : conception de requêtes dynamiques optimisées.

Contexte & Environnement

- **Plateforme SaaS knowledge-first** : dédiée aux équipes data, R&D et produit.
- **Sources hétérogènes** : PDF, vidéos, extraits de code, bases scientifiques.
- **Environnement multi-agents** : usage collaboratif, multi-utilisateurs.

Objectifs Clés

- **Structuration dynamique** : construction et mise à jour continue du graph.
- **Personnalisation contextuelle** : agents spécialisés selon profil utilisateur.
- **Automatisation complète** : ingestion, versioning, enrichissement.
- **Accessibilité fluide** : agents déployés via API, UI, ou messagerie.

Technologies & Infrastructure

- **Langages & frameworks** : Python, PyTorch (Autograd, TorchScript, PyTorch Lightning), LangChain, Hugging Face.
- **Bases & indexation** : Neo4j (graphes), Elasticsearch (vector search).
- **MLOps & CI/CD** : MLflow, GitHub Actions, Docker.
- **Orchestration** : Airflow pour les DAGs, N8N pour les automatisations.

DECATHLON | OCTO TECHNOLOGY – SUPPLY CHAIN FORECAST

Lead senior Data scientist – Forecast Supply Chain & R&D

Mission longue / Paris, France.

Mission

- **Prévision de la demande** : Développement de modèles prédictifs, incluant des réseaux RNN et LSTM pour capturer les dynamiques temporelles complexes de la demande.
- **Optimisation des stocks & réapprovisionnements** : Réduction des ruptures et surstocks.
Industrialisation : Automatisation des pipelines, intégration continue, monitoring des modèles.

Contexte & Environnement

- **Secteur** : Grande distribution, e-commerce, retail omnicanal.
- **Complexité de la supply chain** multi-entrepôts et multi-canaux.
- **Comportements consommateurs** dynamiques et saisonnalité marquée.
- **Contraintes logistiques fortes** (délais, capacité, coûts).
- **Équipe projet** : 1 PManager / 4 Data Scientists / 2 Data Analysts / 1 ML Engineer

Objectifs Clés

- **Forecast multi-échelles** : Prévisions du produit au stock, du jour à la saison.
- **Prévisions probabilistes** : Distributions, quantiles, intervalles de confiance.
- **Apprentissage incrémental** : Architecture par cutoff, mise à jour continue des modèles.
- **Automatisation** : Orchestration complète des pipelines de A à Z.
- **Optimisation logistique** : Diminution des ruptures, ajustement dynamique réapprovisionnement.

Technologies & Infrastructure

- **Langages & Frameworks** : Langages & Frameworks : Python (PySpark, scikit-learn, TensorFlow avec descente de gradient et backpropagation), FastAPI, Streamlit, Poetry
- **Cloud & Orchestration** : Databricks, AWS S3, Airflow, Databricks JOBS Templates
- **Stockage & Pipelines** : Colibra, Parquet, Delta Lake
- **CI/CD & MLOps** : CI/CD & MLOps : GitHub Actions, Docker, MLflow, Git ; gestion du surapprentissage (overfitting) via régularisation par Dropout et Batch Normalization
- **Monitoring & Visualisation** : Tableau, dashboards Databricks

SKETCHNGO | AI-POWERED BUSINESS ASSISTANT

Lead AI Product Engineer – Recommendation Systems & Agent

Mission longue / Paris, France.

Mission

- **Développement d'un système de recommandation intelligent** : assistance personnalisée pour chefs d'entreprise.
- **Conception et orchestration de parcours utilisateur guidés par IA.**
- **Génération de prompts dynamiques**, concaténation et contextualisation via RAG, avec encodage sémantique par auto-encodeur variationnel (VAE)..
- **Développement du front (React) et du back-end (Symfony)**, intégration temps réel (messaging, navigation).
- **Création d'agents spécialisés** fondés sur différentes bases de connaissances (business plan, SWOT, RACI, PESTEL, prévisionnels).
- Appel d'API ChatGPT pour la génération automatisée de livrables à forte valeur ajoutée.

Contexte & Environnement

- **Application SaaS d'assistance entrepreneuriale** : pensée pour accélérer la prise de décision et automatiser les tâches critiques.
- **Orientation produit & design** : importance forte de l'UX, de la recherche utilisateur et du prototypage (livrables adaptés).
- **Collaboration multidisciplinaire** : équipe réunissant produit, design, IA et développement.
- Interaction multi-agent : agents experts alignés sur des besoins métier spécifiques.

Objectifs Clés

- **Personnalisation poussée** : recommandations ajustées au profil et objectifs de chaque utilisateur.
- **Automatisation complète** : génération, enrichissement et stockage des livrables (Postgres).
- **Navigation et expérience utilisateur fluide** : UI réactive, parcours intuitifs, messaging en temps réel.
- **Livraison rapide de MVP** : itérations courtes basées sur les retours utilisateurs et la recherche terrain.

Technologies & Infrastructure

- **Front-end** : ReactJS.
- **Back-end** : Symfony (API REST, orchestration logique applicative).
- **IA & agents** : IA & agents : RAG, Langchain, API ChatGPT, intégrant modules Transformer fine-tunés et logique d'embedding.
- **Base de données** : PostgreSQL (stockage connaissances et livrables).
- **Temps réel** : WebSockets / Messaging API.
- **Méthodologie** : Product Discovery, Product Design Sprint, UX Research.

Senior Lead Data Scientist – Real Estate Forecast & R&D

Mission longue / Paris, France.

Mission

- **ML** : Prédiction d'indicateurs immobiliers via modèles de statistiques géospatiales et de machine learning, avec tests de CNN 2D pour l'analyse de cartes rasterisées et intégration d'embedding géographiques.
- **DE & DA** : Conception de pipelines de données (scraping, ETL), pré-traitement / EDA.
- **R&D** : Cadrage scientifique des problématiques (hypothèses, protocoles), synthèse des modèles.
- **KM** : Construction d'un état de l'art (statistiques spatiales, séries temporelles, ML immobilier) et centralisation des savoirs pour réutilisation.

Contexte

- **Secteur** : Immobilier résidentiel et commercial, forte variabilité géographique et saisonnière.
- **Enjeux** : anticiper les tendances de marché pour guider les décisions d'investissement, intégrer des données hétérogènes (open data, annonces, indicateurs macro).

Objectifs

- **Précision** : atteindre $RMSE < 0,1$ sur indicateurs clés, $R^2 > 0,8$.
- **Robustesse** : modèles probabilistes (intervalles de confiance 80 %) pour mesurer l'incertitude.
- **Efficience** : pipelines automatisés, déploiement continu.

Technologies & Infra

- **Langages** : Python (pandas, numpy, geopandas, scikit-learn, statsmodels)
- **ML & TS** : XGBoost, LightGBM, SARIMA, Prophet
- **Pipelines & Stockage** : Airflow, Docker, SQL/PostgreSQL & PostGIS / AWS S3
- **MLOps** : MLflow, GitHub Actions, Docker
- **Viz** : Tableau, Plotly

Professeur en Mathématiques – Master Data Science & Intelligence Artificielle

3 années universitaires / Lille, France.

Mission

- **Enseignement de niveau master** : conception et animation de cours couvrant les fondements mathématiques et probabilistes pour la data science, l'IA et le machine learning.
- **Structuration pédagogique** : création et mise à jour d'un programme équilibré entre théorie, applications avancées et travaux pratiques.
- **Accompagnement des étudiants** : suivi personnalisé, encadrement de projets, préparation à la recherche ou l'industrie.
- **Mise en œuvre de projets data-driven** : application concrète des concepts via études de cas et analyse de jeux de données réels.

Contexte & Environnement

- **Université Catholique de Lille** : formation en data science, IA et machine learning pour étudiants de master.
- **Public** : étudiants issus de cursus mathématiques, informatiques ou ingénierie, se spécialisant dans l'IA.
- **Environnement collaboratif** : cours en présentiel et/ou hybride, ateliers pratiques, utilisation d'outils numériques modernes, interaction multi-disciplinary.

Programme & Objectifs Clés

- **Fondations** : Outils mathématiques de base – calcul, logique, réseaux de neurones (Perceptron, MLP), épistémologie...
- **Théorie des systèmes** : introduction aux systèmes dynamiques & adaptatifs complexes.
- **Probabilités avancées** : mesure, distributions, statistiques asymptotiques, processus stochastiques, séries temporelles, géométrie de l'information, modèles RNN/LSTM, réseaux bayésiens.
- **Observation et analyse des données** : statistiques descriptives, exploration, visualisation.
- **Inférence et estimation** : estimation de paramètres, conception expérimentale, tests d'hypothèses, arbres de décision, modèles bayésiens.
- **Modélisation statistique** : régressions linéaires (simples, multiples), méthodes alternatives, tests paramétriques et non paramétriques.
- **Modélisation non linéaire** : modèles graphiques probabilistes, réseaux, théorie de la percolation, valeurs extrêmes, modélisation basée agents, statistiques spatiales.
- **Projets pratiques** : études de cas, analyse de grands jeux de données, préparation à l'environnement professionnel.

Technologies & Outils

- **Langages & plateformes** : Python, R, MATLAB, LaTeX, SAS, SPSS, SQL (avec Autograd, DataLoader, TorchScript), R, MATLAB...
- **Environnements de travail** : Jupyter Notebooks, Tableau, Power BI, Excel, GitHub.
- **Bibliothèques IA & data science** : NumPy, Pandas, SciPy, Matplotlib, TensorFlow, PyTorch.

Modalités pédagogiques

- Cours magistraux, travaux dirigés, projets tutorés, évaluations continues, ressources interactives, encadrement personnalisé.

Research Scientist – AI / Machine Learning & Data Science Workflows

Expérience de fondateur / Lille & Paris, France.

Mission

- **Développement logiciel open-science** : création d'une solution de cartographie des connaissances dédiée à l'organisation et à la visualisation avancée des savoirs issus de la recherche académique.
- **Optimisation des workflows scientifiques** : automatisation de la gestion bibliographique, rédaction d'états de l'art, géolocalisation et structuration des ressources à travers des pipelines de data science robustes.
- **Gestion et valorisation de la connaissance** : conception de processus pour la documentation, publication et communication des résultats auprès de la communauté scientifique, favorisant la collaboration et la diffusion.
- **Application des sciences des données** : modélisation expérimentale, analyse statistique avancée, et développement d'outils reproductibles pour la gestion des données de recherche.
- **Collaboration SaaS-Université** : pilotage de synergies entre exigences scientifiques et efficacité entrepreneuriale, intégration d'innovations techniques dans une infrastructure SaaS scalable.

Contexte & Environnement

- **Partenariat stratégique** : alliance entre une startup deeptech (WEEKI.IO) et l'Université de Lille.
- **Ouverture et interopérabilité** : adoption de méthodologies open-science, intégration avec les pratiques et outils de la communauté académique et industrielle.
- **Sources multiples et hétérogènes** : articles, bases de données expérimentales, scripts de codes, documents disciplinaires.
- **Équipe pluridisciplinaire** : chercheurs, ingénieurs data science, développeurs software, managers produit et experts AI.

Objectifs Clés

- **Structuration intuitive des connaissances** : génération cartographique dynamique par thématique, discipline ou méthodologie.
- **Automatisation du cycle recherche-publication** : des premiers états de l'art à la valorisation et la dissémination des résultats.
- **Suivi, traçabilité et reproductibilité** : stockage systématique des scripts, analyses, référencements et versions de données.
- **Valorisation académique et entrepreneuriale** : mise en place d'outils de diffusion, de reporting et de communication multidirectionnels.

Technologies & Stack

- **Recherche & Publication** : Zotero, Mendeley, EndNote, LaTeX, Microsoft Word, ORCID, Google Scholar, Microsoft SharePoint.
- **Data Science & Analyse** : Python, R, MATLAB, SPSS, Jupyter Notebooks, GraphPad Prism.
- **Datavisualisation** : WebGL, Tangram, Maplibre, Leaflet, Tableau, Power BI.
- **Dev & Collaboration** : Git, GitHub, Monday.com, Slack.
- **Infrastructure SaaS & Business** : React, Next.js, PHP Symfony, AWS, Kubernetes, Docker, MySQL, PostGIS, QuickBooks, SQL, JIRA, Excel, TileServer.

Scientific Lead & Data Scientist – Extraction, Sélection et Analyse des Articles Scientifiques
Mission longue / Lille, Loos & Remote, France.

Mission :

Développement d'un algorithme d'extraction et de sélection d'articles scientifiques dédié à l'évaluation clinique des dispositifs médicaux

- Conception d'un pipeline automatisé combinant web scraping, machine learning et visualisation interactive pour accélérer la constitution des rapports d'évaluation clinique (marquage CE).
- Recueil automatisé des références scientifiques : extraction, classement et enrichissement des métadonnées depuis PubMed, Google Scholar, ScienceDirect, Embase.

Optimisation du process de revue systématique

- Classification hybride (humain + ML) des études pertinentes selon des critères d'inclusion/exclusion définis (PICO/PICOS, PRISMA...).
- Intégration d'un scoring automatisé, module de ranking, et possibilité de sélection manuelle pour garantir qualité et transparence du process réglementaire.

Génération de tableaux de bord analytiques et reporting dynamique

- Conception d'une interface SaaS en Python/Dash pour la visualisation, le suivi et l'auditabilité de l'ensemble de la démarche, conforme aux exigences réglementaires (MEDDEV 2.7/1, MDCG...).
- Export multi-formats : BibTeX, CSV, XLSX, au service des utilisateurs métiers (affaires réglementaires, consultants, fabricants DM).
- Recherche & Développement en intelligence artificielle appliquée au réglementaire santé
- Implémentation de méthodes avancées de Natural Language Processing (NLP) pour l'extraction d'informations structurées sur la sécurité et la performance des dispositifs.
- Approche modulaire : prototypage, tests itératifs, évaluation comparative avec l'état de l'art (Abstrackr, Rayyan, ASReview...).

Contexte & Environnement

- **Collaboration** : Start-up deeptech (WEEKI) + cabinet expert Affaires réglementaires (B CONSULTANT), appui d'une équipe universitaire multidisciplinaire.
- **Approche "human-in-the-loop"** : L'utilisateur garde la main sur la décision, l'algorithme assiste et explicite chaque étape, favorisant l'appropriation et la confiance.
- **Interopérabilité** : Compatibilité avec les référentiels santé (PRISMA, PROSPERO), architecture SaaS scalable (microservices).

Objectifs Clés

- **Réduction drastique du temps de revue de littérature réglementaire** (de plusieurs semaines à quelques heures).
- **Renforcement de la reproductibilité, traçabilité et transparence des processus d'évaluation clinique.**
- **Facilitation de la montée en compétences IA des professionnels Affaires réglementaires.**
- Industrialisation des workflows de production documentaire réglementaire (CE/EU-MDR).

Technologies & Stack

- **Data & Extraction / Web Scraping** : Python (Beautiful Soup, Selenium, Scrapy), Requests.
- **Machine Learning & Analysis** : Scikit-Learn, Pandas, Numpy, TensorFlow.
- **NLP & Text Mining** : SpaCy, NLTK, Transformer BERT avec mécanismes d'attention, Flair.
- **Datavisualisation & Frontend** : Dash (Plotly), ReactJS.
- **Infra & SaaS** : API REST, Docker, Heroku/Cloud, Stripe (paiement), gestion des tokens/authentication.
- **Outils réglementaires** : BibTeX, PRISMA, PICOS, SLR Toolbox.

PhD Student – Probabilistic Graphical Models & Machine Learning

24 mois / Paris, France.

Mission

- **Comparaison de modèles graphiques probabilistes** : évaluation des approches bayésiennes dirigées face aux modèles non dirigés (Markov).
- **Benchmark avancé** : mise en œuvre et comparaison de réseaux bayésiens et de réseaux de Markov pour la modélisation des menaces cyber.
- **Transition stratégique** : choix des modèles non dirigés pour explorer leur flexibilité face à des scénarios complexes de cybersécurité.
- **Exploration dynamique** : analyse approfondie des réseaux de Markov pour détecter, anticiper et modéliser les risques cyber émergents.

Contexte & Environnement

- **Recherche académique appliquée** : positionnée à l'interface de la sécurité, du machine learning et des statistiques avancées.
- **Enjeux cybersécurité** : anticipation des menaces, amélioration de la détection et de la résilience des systèmes.
- **Volumes de données hétérogènes** : logs réseau, incidents, historiques d'alertes de sécurité.
- **Intégration cloud** : exploitation étendue de la suite Azure pour l'analyse, le stockage et l'orchestration de données.

Objectifs Clés

- **Comparer et valider** : évaluer la pertinence des modèles dirigés vs non dirigés selon différents cas d'usage cyber.
- **Optimisation** : affiner les méthodes pour un diagnostic plus précis des attaques et vulnérabilités.
- **Scalabilité & automatisation** : traitements massifs via Azure Data Factory et automatisation des flux analytiques.
- **Valorisation scientifique** : contribution à la littérature, présentations en conférences spécialisées.

Technologies & Infrastructure

- **Langages & frameworks** : Python, R, TensorFlow, NumPy, Pandas, Matplotlib.
- **Modélisation** : Dynamical Bayesian Networks, Markov Networks couplés avec architectures VAE pour inférence latente.
- **Cloud & Data** : Azure Machine Learning, Azure Security Center, Azure Data Factory, Azure SQL Database.

Research Scientist ML / AI / Predictive Modeling

12 mois / Lille, France.

Mission

- **Modélisation prédictive** : création de modèles ML pour l'analyse de comportements physiques et physiologiques humains (notamment via CNN 1D et RNN).
- **Traitement de données de capture de mouvement** : extraction et transformation de signaux pour analyse comportementale.
- **Projections riemanniennes** : application de techniques géométriques avancées pour représenter la dynamique du mouvement.
- **Analyse en contexte social** : identification des patterns de comportements et réponses physiques en situations variées.

Contexte & Environnement

- **Projet interdisciplinaire** : croisement entre sciences du comportement, IA et neurosciences sociales.
- **Technologies de capture** : systèmes de Motion Capture et OpenPose pour l'enregistrement de mouvements humains.
- **Collaboration scientifique** : travail en équipe avec chercheurs data, psychologues et physiologistes.

Objectifs Clés

- **Compréhension fine** : modéliser précisément les dynamiques du mouvement et des états physiologiques.
- **Génération d'insights** : produire des analyses quantitatives sur les interactions et réponses sociales humaines.
- **Validation scientifique** : assurer la robustesse des modèles par des expérimentations et validations croisées.
- **Contribution académique** : enrichir les connaissances en interaction humaine via des publications et présentations.

Technologies & Infrastructure

- **Langages & frameworks** : Python, R, MATLAB, TensorFlow, Keras
- **Capture & traitement** : Motion Capture Systems, OpenPose
- **Modélisation avancée** : Riemannian Geometry, NumPy, Pandas, Scikit-learn
- **Analyse & visualisation** : Matplotlib, Seaborn (si applicable)