# Théo VERSACE

Développeur IA / Data Scientist

Balma (31130) · 06 98 50 64 51 · theo.versace@proton.me

Github · LinkedIn

## PROFIL

Développeur IA et Data Scientist en dernière année de Master Expert IA & Data Sciences, en alternance depuis 2024 chez un éditeur de logiciels de santé. J'interviens sur l'ensemble du cycle d'un projet d'IA : analyse statistique, modélisation prédictive, deep learning, mise en production et gouvernance des données. Également expérimenté en IA générative (RAG, agents) et en Business Intelligence.

## COMPÉTENCES

**Langages :** Python, SQL (T-SQL, PostgreSQL), C#, R, JavaScript, TypeScript
**Machine Learning :** Scikit-learn, XGBoost, LightGBM, CatBoost, Gradient Boosting, Random Forest, SVM, Optuna, SHAP
**Deep Learning :** PyTorch, TensorFlow / Keras, LSTM, CNN et transfer learning, fine-tuning de modèles de langage
**Data & statistiques :** Pandas, NumPy, SciPy, Statsmodels, tests d'hypothèses, feature engineering, détection d'anomalies
**IA générative :** RAG, LLM, LLM locaux (Ollama), agents IA (Agno), pgvector, Model Context Protocol (MCP), GitHub Copilot
**Mise en production :** FastAPI, Flask, Docker (Compose), CI/CD GitHub Actions, MLflow, Evidently, Prometheus, Loki, Grafana, Git, Jira
**Business Intelligence :** Power BI (DAX), PBIRS, Metabase, SQL Server
**Web :** Next.js, React, Supabase, Vercel
**Sécurité & conformité :** WAF ModSecurity (OWASP CRS), RGPD, HDS, AI Act (UE 2024/1689)

## EXPÉRIENCE

**Développeur IA & Data Scientist, Atout Majeur Concept (alternance)** 2024 – aujourd'hui

*Éditeur de logiciels pour établissements de santé, progiciel Sésame — Toulouse*

**Projet européen Harmonic AI : prédiction de la durée de séjour hospitalier**

- Conception d'un modèle de prédiction de la durée de séjour des patients admis aux urgences, dans le cadre d'un projet européen réunissant huit pays autour de l'IA explicable en santé.
- Analyse statistique de 7 700 séjours du CHU de la Martinique : nettoyage, détection d'hétéroscédasticité par test de Breusch-Pagan, bascule vers des tests robustes (Welch, Kruskal-Wallis, Games-Howell) et transformation logarithmique de la variable cible.
- Feature engineering (temps d'attente aux urgences, créneau d'admission, week-end) et sélection des variables par matrice de corrélation et VIF.
- Benchmark d'une quinzaine d'algorithmes en classification et en régression, optimisation par Optuna, validation croisée et explicabilité par analyse SHAP.
- Mission d'un mois à l'Université d'Évora, au Portugal, pour un cas d'usage à dimension internationale.

**IA générative et outillage des équipes**

- Conception d'un système RAG interne indexant le code et la documentation technique (ingestion, chunking, embeddings, base vectorielle), afin d'obtenir des réponses contextualisées sans exposer le code source à l'extérieur.
- Intégration du Model Context Protocol dans la chaîne de développement, dont le serveur Figma pour générer des composants front-end à partir des maquettes validées.
- Déploiement de GitHub Copilot dans l'entreprise et formation des équipes de développement.

**Business Intelligence**

- Déploiement de Power BI et PBIRS chez plusieurs hôpitaux clients : rapports Gestion des lits en temps réel et suivi des mouvements patients multi-établissements.
- Mesures DAX avancées, optimisation des sources par création de vues SQL Server.
- Migration de rapports Power BI vers Metabase (modélisation en étoile, requêtes T-SQL), en environnement HDS et RGPD.

**Professeur particulier en mathématiques** 2023 – 2024

## PROJETS

**VéloCast Toulouse, prévision de disponibilité des vélos en libre-service** 2026

*Projet personnel*

- Service de prévision du nombre de vélos disponibles dans les ~450 stations VélôToulouse à 15, 30 et 60 minutes, à partir de données collectées toutes les 5 minutes (API JCDecaux, Open-Meteo).
- Modèle LightGBM réentraîné chaque semaine sur 2 ans d'historique. Testé sur 6 mois de données jamais vues : erreur moyenne de 1,59 vélo à 60 minutes, contre 1,79 pour une prévision naïve (nombre de vélos supposé inchangé), soit 11 % de moins.
- MLOps sans serveur : réentraînement automatique, détection de dérive (Evidently), alertes, nouveau modèle affiché seulement s'il bat les références en conditions réelles. API FastAPI, PostgreSQL, site Next.js avec carte interactive.

**Resilience-Alpha, banque fictive qui détecte une cyberattaque et se répare seule** 2026

*Projet personnel*

- Système de paiement complet sous Docker Compose (React, FastAPI, PostgreSQL), derrière un WAF ModSecurity + OWASP CRS en mode bloquant, supervisé par Prometheus, Loki et Grafana.
- Détection en direct de cinq attaques (brute-force, flood, scan, injection, jeton forgé) à partir des logs : un modèle d'anomalie appris sur le trafic normal seul, complété par un classifieur fine-tuné.
- Remédiation automatique en quelques secondes : IP bannie au WAF, services touchés redémarrés, puis un vrai virement de test qui prouve la reprise.
- Rapport d'incident rédigé par un LLM local (Ollama) et versé dans une base RAG pgvector ; une équipe d'agents Agno propose une solution, qu'un humain valide.

**MECHA.IA, maintenance prédictive industrielle** 2026

*Projet de certification (MSPR)*

- Deux modèles PyTorch complémentaires : un réseau dense pour la détection instantanée de panne, et un LSTM bidirectionnel pour estimer le temps restant avant défaillance (RUL).
- Pipeline ETL avec anonymisation SHA-256 dès l'ingestion et génération de 717 000 lignes synthétiques. Mise en production via une API FastAPI conteneurisée sous Docker, avec CI/CD GitHub Actions.
- Apprentissage continu par Replay Buffer pour prévenir l'oubli catastrophique, alimenté par une boucle de feedback intégrée à un dashboard React. Classification du système au regard de l'AI Act.

**WildLens, classification d'empreintes animales** 2025

*Projet de certification*

- Identification de l'espèce à l'origine d'une empreinte parmi 13 mammifères, pour une application de sensibilisation à la faune sauvage.
- Transfer learning sur MobileNetV2 et EfficientNetB0, avec entraînement en deux phases et fine-tuning des dernières couches à learning rate réduit.
- Augmentation de données et filtrage pour compenser un jeu de 252 images déséquilibré : 87,2 % d'accuracy et 96,4 % de top-3, avec analyse des confusions inter-espèces.
- Déploiement via une API REST Flask et une interface web responsive.

**PolePickF1, application de pronostics Formule 1** 2025

- Application développée en Next.js avec Supabase et déployée sur Vercel, disponible en version web et en application mobile publiée sur l'App Store.

**Autres** 2023 – 2025

- Jeux mobiles (Skyjo, Uno). Programme de repas éco-responsables avec les cantines de l'INSA Lyon.

## FORMATION

**Master Expert IA et Data Sciences, EPSI Toulouse** 2025 – 2027

**Bachelor Développeur IA et Data Analyst, EPSI Toulouse** 2024 – 2025

**INSA de Lyon, département Bio-informatique et Modélisation** 2022 – 2024

Deux années validées, réorientation choisie vers l'IA et la data science.

**CPGE, lycée Jean Mermoz** 2020 – 2022

**Baccalauréat, lycée Déodat de Séverac** 2020

## CERTIFICATIONS

- Learning AI Through Visualization, Columbia University (juin 2025).
- Anthropic Academy (mars 2026) : AI Fluency, Claude with the Anthropic API, Claude Code in Action, Introduction to MCP, MCP Advanced Topics.

## LANGUES

Français langue maternelle. Anglais niveau B2, pratiqué en contexte professionnel international (projet Harmonic AI, mission au Portugal).

## DIVERS

- Permis B
- Passions : Football, Formule 1, Triathlon, Jeux Vidéo