

## David TCHATCHOUA

AI Engineer (7 years of experience)

Python Expert

💡 "Innovation distinguishes between a leader and a follower." – Steve Jobs

Ambitious and applied AI Engineer, **Harvard graduate** in Machine Learning, **Microsoft Certified DP 100 and DP 102**, with extensive expertise in deploying ML, DL and LLMs applications in AWS, GCP, Azure, Hugging Face, I also have solid skills in **DevOps, API, SQL & NoSQL databases**. Proficient in **Python** (PyTorch, TensorFlow, Scikit-Learn, NumPy, Pandas, Flask, etc.) **Rust**, and **Front-end** solutions (NodeJS, React, Vue JS, Typescript, JavaScript).

**Founder of IA Pour Tous** – Initiative to democratize AI by showcasing **hands-on professional use cases** and **bridging the gap between AI research and real-world business impact**.

## Professional experience

Since Jan 2025:

**AI Engineer at Frenchtechacademie**

**Building Agents AI with LangGraph (LangChain, Langsmith), CrewAI, AutoGen, MetaGPT, LlamaIndex and Smolagents**

- ✓ Web Agents with Memory, Human in the loop
- ✓ SQL Agents
- ✓ Data Analysis Agents

Nov – Dec 2024:

- ✓ **Automating analysis with AI of CVs for HR** (OpenAI API, PDF extraction, CSV and SQL storage, Dashboard, Blob File, Code optimization, Scrum Agility Project management).



✉ [davidtchouta@gmail.com](mailto:davidtchouta@gmail.com)

🏠 Address: **1 B Rue Sadi Carnot  
74100 Ville-La-Grand – France**

🎂 **32 years old**

🔍 <https://www.linkedin.com/in/david-tchatchoua-a16a89123/>

☎ **+41 78 310 57 62**

☎ **+33 7 82 99 45 61**

## About me

Personal Page 📄

<https://frenchtechacademie.fr/tchatchoua>

YouTube ▶

<https://www.youtube.com/@davidtchatchoua>

GitHub 📁

<https://github.com/davidtchouta>

## Languages

**English:** Fluent 🗨️

**German:** Intermediate 📖

**French:** Mother tongue 🏠

## Skills

- ✓ LLMs (GPT, Encoders, Transformers, Decoders) : Llama2, BERT, GPT 2, Mistral, etc.

- ✓ **Financial News monitoring and analysis with AI** (Yahoo Finance API, Sentiment Analysis, Financial Portfolios Management, OpenAI API, PDF extraction, CSV and SQL storage, Data Labeling, Dashboard, Code optimization, Scrum Agility Project management).
- ✓ **Write, Read and Investigate contracts with AI to find hidden statements** (OpenAI API, PDF Extraction, Python Clean Code)
- ✓ **AI App for Financial Support Engineering** (SQL database, similarity Scrum Agility Project management).
- ✓ **AI App for Financial Support Engineering** (SQL database, similarity search, python clean code, OpenAI API, Tokens Management).
- ✓ **Financial Technical Analysis with AI** (MSI, Moving Average, Momentum, Python clean code, OpenAI API, Parabolic stop and Reverse, Ichi Moku Cloud, RSI, Rate of change, Bollinger Bands, Fibonacci Retracement, etc.).
- ✓ **Model optimization in Regression Problem** (stacking models, Ensemble models, Boosting Models: LightGBM, XGBoost, Deep Learning Models, Auto Gluon, Features Analysis, EDA, etc.).
- ✓ **Microsoft Copilot Creation Agent Tool:** Set a Microsoft Copilot Agent Tool for Human Resource Department.

**Jul-Nov 2024:**

**AI Engineer at Frenchtechacademie**

**Project LLM (JHK303): RAG, Fine tuning, Prompt Engineering**

- ✓ Implemented Retrieval-Augmented Generation (RAG) using **LangChain** and **Qdrant**, improving document retrieval accuracy, and integrated **FAISS** for efficient

- ✓ Cloud Deployment (AWS, GCP, AZURE)
- ✓ Model Building & Deploy (AWS SageMaker, Vertex AI, Azure ML, Hugging Face)
- ✓ ML Pipelines: MLFlow, ZenML, PyCaret, DVC, Kubeflow, TFX
- ✓ Programming languages (Python, Scala, PySpark, R, SQL, Java, Typescript, React, NodeJS, Swift, Objective-C++, C++, Rust)
- ✓ Tools and libraries (TensorFlow, PyTorch, GPU, TPU, JAX, Keras, Caret, Pandas, etc.)
- ✓ CI/CD: GitHub Actions, Azure DevOps, CloudWatch, Gitlab
- ✓ IAC: Terraform, Ansible, Vagrant, Parker
- ✓ Security: Hashi Corp Key Vault, Azure Key Vault, GitHub Secrets, Firewall, RBAC, VPC, VPN
- ✓ Virtualization, Docker, Kubernetes
- ✓ Prometheus, Grafana
- ✓ Spark, Hadoop, Snowflake Apache Kafka
- ✓ Databricks, Dataiku, MLRun
- ✓ VCS: Git
- ✓ Neural Networks, Graphs NN, Bayesian Networks, Markov Chains/HMM, Probabilistic Graphical Models, Reinforcement Learning
- ✓ AI Agents (LangGraph, CrewAI, AutoGen, Pydantic, LlamaIndex, AutoGPT, Langflow)
- ✓ Broker : RabbitMQ, Celery

similarity searches.

- ✓ Fine-tuned **Hugging Face Transformers** for specialized tasks using **LoRA** to reduce resource usage while maintaining performance.
- ✓ Managed experiments and model versions with **MLFlow**, and shared models via **Hugging Face Model Hub**. Optimized inference with **ONNX**, deployed on **Kubernetes for scalable production**.
- ✓ Developed advanced **Prompt Engineering** strategies with **OpenAI API and Anthropic Claude** to maximize response relevance.
- ✓ Monitored model performance with **Prometheus and Grafana**, setting up proactive alerts.
- ✓ Automated retraining and deployment pipelines using **Azure Machine Learning for continuous model updates**.
- ✓

#### **MLOps Project: Deploying an ML Application for Insurance Premium Predictions on AWS, Azure, GCP**

- ✓ **Developed an insurance premium prediction model** using **Python** and integrated it into a **user-friendly interface with Flask**.
- ✓ Registering **experiment tracking, model registry, metrics monitoring with MLFlow**
- ✓ **Virtualized and containerized the application** using **Docker**, ensuring reproducibility across **AWS, Azure, and GCP environments**.
- ✓ Managed **cloud infrastructure with AWS Fargate, Google Kubernetes Engine (GKE), and Amazon ECS for scalable, efficient container orchestration**.
- ✓ Set up **container registries on AWS ECR, GCR, and ACR for secure and traceable deployment**.

- ✓ Implemented **A/B Deploy strategies** on **Azure Web App** and **Amazon SageMaker for continuous model validation**.
- ✓ **Monitored application performance** with **AWS CloudWatch, GCP Monitoring, and Azure Monitor**, optimizing key metrics and detecting anomalies.
- ✓ **Secured data and compliance** through **advanced firewall configurations** across cloud platforms.
- ✓ Used **Pickle for model serialization** and **Amazon S3 for efficient data storage and version control**.

#### **Projects Links:**

- ✓ [Deploy ML flask App in AWS Fargate solution](#)
- ✓ [Deploy ML flask App in Google Cloud Platform \(GCP\)](#)
- ✓ [Deploy ML flask App in Azure Cloud](#)
- ✓ [Deploy Model in AWS SageMaker](#)

**Jan-Jun 2024:**

**AI Engineer at Frenchtechacademie**

#### **Computer Vision Project:**

- ✓ Implemented **deep learning solutions for computer vision tasks**, including **object detection and real-time object tracking**, with strong integration of **PyTorch**.
- ✓ Designed a **data processing pipeline to train model**.

- ✓ **Keep up to date** the **latest advancements** in the field
- ✓ **Models** used:
  - **Object detection:** Table-transformer-detection (Microsoft), Detr-resnet (Facebook), Yolo (Ultralytics)
  - **Image Segmentation:** SegFormer, Mask2former, Clipseg-rd64

### ML Projects:

- ✓ Gradio Application for Iris Classification
- ✓ Developed an **interactive ML-powered application** using **Gradio** and **deployed it on Hugging Face**.
- ✓ **Designed the solution to classify** the Iris dataset using a PyCaret-based workflow.
- ✓ **Preprocessed data** and **selected the optimal model** (e.g., Random Forest, Decision Tree) using PyCaret's compare model's method for **accuracy optimization**.
- ✓ Implemented a **robust prediction interface** with customizable inputs (dropdowns and sliders) for **user-friendly interaction**.
- ✓ Integrated **real-time predictions** with **probability scores** for different Iris species (setosa, versicolor, virginica).
- ✓ Demonstrated **proficiency** in **Python libraries** like **Pandas, PyCaret, and Gradio** for **machine learning** and **deployment**.
- ✓ **Hosted the app on Hugging Face Spaces**, enabling global accessibility and **showcasing practical AI deployment experience**.
- ✓

### Project link:

- ✓ [Build Gradio app and Deploy in Hugging Face](#)
- ✓ Streamlit App for Health Insurance Cost Prediction
- ✓ Developed an **intuitive machine learning application** using **Streamlit** to estimate **health insurance costs** based on user provided inputs

- ✓ Created an **interactive user interface** with **real-time prediction capabilities**, **allowing both online and batch processing** options for input data.
- ✓ Enabled predictions based on features such as age, gender, BMI, smoker status, number of children, and region.
- ✓ **Deployed the application on Render** using **GitHub** for seamless **version control** and **hosting**, ensuring global accessibility via **live demo**.
- ✓ Integrated **data preprocessing pipelines** to handle **single-entry predictions** and **batch file uploads for scalability**.
- ✓ Utilized **custom branding and visualizations**, including **graphical assets**, to enhance the user experience.
- ✓ Demonstrated **strong expertise** in **Python, Streamlit, PyCaret**, and **deployment to cloud hosting platforms**.

### Other projects:

- ✓ [Build and Train a POC of text generating from scratch](#)
- ✓ [Computer Vision: Object Detection](#)
- ✓ [Test my Email Spam Classifier](#)
- ✓ [How to Build ML Pipelines with PyCaret and MLFlow](#)
- ✓ [Computation of Predictions, Overall Accuracy, Variable Importance on Iris dataset](#)
- ✓ [Computation of RMSE \(Root-Mean-Square Error\) on a Linear Regression Model in R](#)

- ✓ [Does Sample Size have an Impact on RMSE \(Root Mean Square Error\)?](#)
- ✓ [What Happens to RMSE when Dependent and Independent Variables are Highly Correlated?](#)
- ✓ [What Happens to RMSE when the Vector of the Independent Variable Increases?](#)

Jan 2022-May 2024

MLOps & DevOps Engineer at Frenchtechacademie

#### MLOps Projects:

- ✓ Set up **performance monitoring** for machine learning models using **Prometheus** to **collect real-time metrics** (response time, prediction success rates, resource usage), and **visualized them on custom Grafana dashboards**.
- ✓ Automated **monitoring pipelines with GitHub Actions**, enabling **continuous performance testing and automatic model updates** based on **collected metrics**.
- ✓ Developed a **FastAPI RESTful API** to manage and monitor prediction requests, collecting **performance data** at every interaction.
- ✓ Built an **interactive user interface with Streamlit** to allow **Realtime performance analysis and visualization**, integrated with **Grafana dashboards**.
- ✓ Managed **model serialization with Pickle** and **tracked experiments using MLFlow** to **log key performance metrics and version control**.
- ✓ Implemented a **robust data storage solution** with **Azure Blob Storage**, automating data archiving and model versioning to ensure data integrity and easy access.

#### DevOps Projects:

- ✓ **Automated CI/CD pipelines** using **GitHub Actions, GitLab and Jenkins** to streamline the continuous deployment of machine learning models and applications.
- ✓ Secured **sensitive information** with **GitHub Secrets** and implemented automated tests to validate model integrity after each update.
- ✓ Leveraged **Terraform for cloud infrastructure management, automating resource provisioning, and using Ansible for consistent server configuration**.
- ✓ Created **replicable development environments** with **Packer and Vagrant** to ensure standardized setups across teams.
- ✓ **Monitored infrastructure and model performance** with **Prometheus, Grafana, and ELK Stack**, providing real-time insights and proactive issue resolution.
- ✓ **Secured API keys and credentials** using **Vault by Hashi Corp**, automating **secret management and access policies**.
- ✓ Integrated **SonarQube** to ensure **code quality and security** across ML pipelines.

- ✓ **Deployed ML models using Azure DevOps and GitLab CI/CD**, ensuring **scalability** and **flexibility** across **multi-Cloud environments** (Azure, AWS, GCP).

**Jan 2021-Dec 2021**

**Data Scientist at TotalEnergies**

- ✓ Developed a comprehensive **data analysis workflow using both R and Python**, ensuring efficient data processing and insights extraction.
- ✓ Conducted a **comparative analysis** between **Power BI** and **Tibco Spotfire** to evaluate their **performance and suitability** for business intelligence tasks.
- ✓ Worked in an **Agile environment**, collaborating with **cross-functional teams** to ensure **timely delivery of high-quality solutions**.
- ✓ Integrated and utilized **REST APIs** for seamless data access and application integration.
- ✓ **Managed tasks and project progress with Jira**, ensuring efficient project tracking.
- ✓ **Built and deployed machine learning models**, driving data-driven decision-making and predictive analytics.

**Sept 2018-Mar 2019**

**Structuring assistant of employee savings and retirement funds at Amundi Asset Management**

- ✓ Developed advanced **VBA-Excel macros** to **automate repetitive tasks and improve workflow efficiency** in data management processes.
- ✓ Utilized **Access-VBA** to **create and manage databases**, enhancing data storage and retrieval operations.
- ✓ **Wrote complex SQL queries** to extract, manipulate, and analyze large datasets, ensuring accurate and **insightful reporting**.

- ✓ **Applied expertise in fund management to analyze financial data**, optimize **portfolio performance**, and support decision-making.
- ✓ Leveraged **data engineering** skills to **build robust data pipelines**, ensuring seamless data flow across systems.
- ✓ Conducted **in-depth data analysis** to uncover trends and provide actionable insights for **business optimization**.

**Apr 2017-Sept 2017**

**Research assistant at INRA-AgroParisTech – LEF**

- ✓ Conducted an analysis to **estimate the risk aversion and time preferences of forest owners**, utilizing advanced simulation techniques to model decision-making under uncertainty.
- ✓ Applied principles from **game theory and microeconomics** to explore strategic interactions between forest owners and other stakeholders, particularly in the context of **climate change**.
- ✓ Analyzed how different levels of **risk aversion influence decisions related to forest conservation and resource management**.
- ✓ Employed **econometric methods** to rigorously quantify these behaviors, using real-world data to validate the models and assess policy implications.

## Interests

- ✓ **Hobbies:** Golf, reading editorials, hiking, gourmet cooking, learning video editing
- ✓ **Staying updated on AI technologies:** Medium blog, FreeCodeCamp, Coursera, Microsoft Learn, Papers with Code, arXiv, etc.
- ✓ **Aspirations:** Learning a new Language, traveling to every continent, contributing to an Open-Source Project, sharing my knowledge.

## Assets

- ✓ **Rigorous**
- ✓ **Autonomous**
- ✓ **Team Player**
- ✓ **Curious**
- ✓ **Problem solving**

## Degrees and Training

**Since Janv 2025:**

**German language exam preparation** / Duolingo  
Ville-la-Grand

**May 2024 – Oct 2024: AI Engineer Certificates –  
Microsoft – Coursera**

[Large Language Model Operations \(LLMOps\)  
Specialization Certificate – Coursera](#)

[Microsoft Certified: Azure AI Engineer Associate \(DP  
102\)](#)

[Microsoft Certified: Azure Data Scientist Associate  
\(DP 100\)](#)

[Machine Learning Operations \(MLOps\)  
Specialization Certificate – Coursera](#)

[Applied Python Data Engineering  
Specialization Certificate – Coursera](#)

[Data Science: Machine Learning  
Certificate - Harvard University School  
of Public Health](#)

**Sept 2018 – Mars 2020:**

**Developer Certificate Java Full Stack**  
(12/2019 - 03/2020) - M2i Formation

**Computer Sciences Certificate**  
(12/2018 - 06/2019) – Sololearn

**Oct 2017-Dec 2018:**

**MBA in Market Finance, Asset  
Management and International  
Commodities Trading** / Paris ESLSCA  
Business School Paris, France: Asset  
Management, Derivatives Products,  
Investment Bank, etc.

**Sept 2014 – Sept 2017 :**

**Economist Engineer** / École  
d'Économie de l'Université Clermont  
Auvergne Clermont-Ferrand, France :  
Econometric models, Statistics,  
Macroeconomic Forecasting, etc.