

Farouk HAKEM

DATA SCIENTIST

📞 +33 07 81 56 13 66 | ✉️ hakem.farouk@gmail.com | 🏠 www.faroukhakem.com | 🐙 github.com/fhakem | 🔗 linkedin.com/in/fhakem

Personal Profile

Data Scientist with more than 5 years of experience. I develop machine learning models and deploy them on an industrial scale. I work transversally on the model lifecycle, and make sure models are properly integrated into complex systems.

Work Experience

Mirakl

Data Scientist

Paris, France

Jan 2025 – Present

- Developed an end-to-end **e-commerce data automation platform** enabling marketplaces to seamlessly onboard sellers regardless of their catalog formats.
- Built **auto-mapping systems** leveraging advanced machine learning models to automatically match heterogeneous product attributes and categories.
- Designed and fine-tuned models for **complex information extraction**, classification into **multi-level taxonomies**, and automated product compliance rewriting.
- Implemented **multimodal extraction pipelines** using **CLIP** and vision-language models to extract product information from images.
- Integrated **translation and rewriting components** to ensure multilingual support and compliance with marketplace requirements.
- Fine-tuned and deployed LLMs such as **OpenAI GPT models** and **Mistral** on a scalable **Databricks** infrastructure to support high-volume catalog processing.
- Designed monitoring and evaluation systems to continuously assess model accuracy, reliability, and latency across pipelines.

LexisNexis

Data Scientist

Paris, France

Jun 2024 – Dec 2024

- Developed and deployed a state-of-the-art **legal chatbot** to provide efficient and personalized legal assistance to clients.
- Fine-tuned large language models **LLM like GPT-4 and Claude** to create highly customized AI assistants tailored to the company's legal domain and customer needs.
- Designed and implemented a robust natural language processing (**NLP**) **pipeline** to understand client queries, classify legal issues, and generate relevant responses.
- Developed an **intent detection system** to effectively allocate tasks and responsibilities among the system various components, ensuring seamless operation and clear accountability.
- Utilized machine learning techniques, such as transfer learning and active learning, to continuously improve the **Retrieval Augmented Generation (RAG)** performance, resulting in a 30% increase in customer satisfaction.
- Designed and implemented **data collection and monitoring** systems to track the chatbot's usage, performance, and user feedback, enabling data-driven optimization.

Shift Technology

Data Scientist

Paris, France

Sept 2021 – May 2024

- Developed and deployed to production **predictive ML models for fraud detection**, increasing savings of concerned insurers clients by 20%.
- Significantly boosted the models accuracy by 70%, using more adapted models, better **data preprocessing and federated learning techniques**.
- Development of a framework allowing data processing and model training for several business lines
- Implementation of **MLOps practices for model management**, particularly for deployment and retraining
- Integrate **ML models into complex systems by creating APIs** that comply with software design standards
- Creation of **cloud environments** correctly sized according to chosen models and data volumes
- Developed a deployment and monitoring system to track the performance of the models and improve them efficiently.
- Facilitating cross-functional meetings to exchange best practices in data science, MLOps, and software design
- Creation of data pipelines using **advanced ETL** tools as well as in-house solutions
- Development of packages to facilitate processing of certain types of data (invoices, ID documents, etc.)
- **Technical Skills:** Python (PyTorch, Pandas, Scikit-learn, LightGBM, XGBoost), C#, SQL, Git, Databricks, Cloud Solutions (Azure, AWS), Terraform, PySpark
- **Soft Skills:** Teamwork, Leadership, Time Management, Client Facing, Communication, Logical Thinking, Presentation Skills

Bouygues Telecom

Machine Learning Intern

Paris, France

March 2021 – Aug 2021

- Trained and deployed to production a Python **NLP engine for classification of intentions and entities** used to create chatbots. This has resulted in savings of 1.2 million per year by replacing the NLP services (IBM Watson and SAP), while increasing the accuracy of the service by 15%.
- Integration with the back-end and implementation of an **advanced CI/CD** (continuous integration/continuous development) process
- Deployment of the solution on Red Hat Openshift / Kubernetes.
- Developed a **recommendation engine** that helps telemarketers improve their customer-facing suggestions, increasing subscriptions by 8%.
- **Technical Skills:** Python (Bert Language Model, NumPy, Pandas, spaCy), Java, AWS (EC2, S3, SageMaker...), Gitlab, Docker, Kubernetes, MLOps, DevOps

Atomic Energy and Alternative Energy Commission (CEA)

Paris

Data Scientist

Aug 2020 - Jan 2021

- Development of a playful interface to raise awareness of the importance of the economic and ecological impact of the computing infrastructure used to **train neural networks on N2D2** (CEA GPU platform). Was presented at CES in Las Vegas in 2021.
- Creating problems requiring AI methods to solve them
- Creation of an API to access model training times and energy costs
- Web technology interface development

Education

Ecole Polytechnique

Paris, France

Master in Data Science

Sept 2020 - Aug 2021

- **Courses:** Data Science, Programming for Data Science, Knowledge Representation and Reasoning, Statistical Theory and Methods, Learning Skills through Case Studies, Artificial Intelligence, Machine Learning, Statistical Learning, Unsupervised learning, Reinforcement Learning, Meta-heuristics, Genetic Algorithms

Telecom SudParis

Paris, France

Generalist Engineer

Sep 2017 - Aug 2020

- Graduated with Distinction.
- **Courses:** Mathematics, Probability, Statistical Analysis, Optimization, Stochastic Processes, Time Series, Telecommunications and Networks

Projects

Computer vision for football

Footbar

- Detection of the different phases of the game, shots/passes and player tracking. Cut the time and cost of annotating videos in half.
- Developed an approach based on the detection of a succession of poses characteristic of the gesture of a shot.
- Utilisation of the audio tracks to develop a classifier to identify important moments, notably using Fourier transforms and ML techniques.
- **Technical Skills:** Python, OpenCV, YOLO, Openpose, Neural Networks, Fast Fourier Transform

Tracking the supply chain of medicines

Chainron

- Development of a decentralized application using blockchain (Ethereum) to monitor the quality of medicines throughout the supply chain.
- Creation of our own smart contracts to validate the passage of a drug from one logistical stage to another
- **Soft Skills:** Entrepreneurship, Innovative spirit, Communication, Curiosity, Self-Learning

Development of a medical chatbot

Tata Consultancy Services

- Development of a chatbot to help people with neurological diseases. Deployed on several platforms (Facebook, Whatsapp, KakaoTalk.)
- Integration of different data sources, both personal (social networks, smartphones, etc.) and open source.
- **Technical Skills:** Dialogflow, Node.js, Rasa, Google Cloud Platform, Firebase, Social Network API

Development of an HR system

IBM

- ERP development using web technologies (Node.js and Vue.js)
- KPI analysis (churn, average rates...)
- Development of an algorithm to predict consultant churn

Skills

Languages	French (mother tongue), English (C1), Spanish (B2), Arabic
Programming	Python (NumPy, Pandas, Tensorflow, Keras, PyTorch, Scikit-learn. etc.), SQL, Machine learning, MLOps, DevOps, Java, C#, OOP, Clean Code, HTML/CSS, JavaScript
Miscellaneous	Linux, Shell (Bash/Zsh), \LaTeX , Tableau, Microsoft Office, Firebase, Git (GitHub/GitLab), Jira, AWS, Azure, GCP, ETL
Soft Skills	Time Management, Teamwork, Problem-solving, Adaptability, Engaging Presentation, Documentation