

# Raphaël Azorin

PhD in Machine Learning with industry experience.  
Looking for exciting challenges as a ML Engineer.

azorin (dot) raphael (at) gmail (dot) com

Born in 1993

French



[raphaaal.github.io](https://raphaaal.github.io)

## EXPERIENCE

**Ubisoft** — Senior Machine Learning Engineer (1.5 year) Sept. 2024 - Present

Bridging the gap between research and product for fraud detection. I develop systems, from proof-of-concept to production, to detect re-offenders among millions of profiles.

**Huawei** — Research Engineer (3 years) June 2021 - June 2024

I led machine learning projects on compression data structures, tabular time-series and constrained networked systems. This yielded 8 publications in top venues available [here](#).

**xCo Analytics** — Data Engineer (1 year) Sept. 2019 - Oct. 2020

I worked on production data pipelines for an analytics platform. I processed terabytes of ticketing and smart building data for data quality and machine learning initiatives.

**Marionnaud** — Data Miner (1 year) Sept. 2018 - Sept. 2019

I worked on customer behavior modeling for product line managers. I developed segmentations, clustering and forecasts to support a launch targeted at 500K customers.

**Markentive** — Marketing Consultant (2.5 years) Feb. 2016 - Sept. 2018

Before venturing into computer science, I was in digital marketing. I helped build the first HubSpot Diamond partner in France, from 5 to 30 employees.

## EDUCATION

**PhD in Machine Learning** — Sorbonne University June 2024

**MSc in Machine Learning** — PSL University (salutatorian) June 2020

**BSc in Computer Science** — PSL University (valedictorian) June 2018

**MSc in Management** — Grenoble School of Business / University of Mannheim June 2015

## PROJECTS

**Dumbo** — Tiny machine learning ([GitHub](#))

ML system for network monitoring. A lightweight classifier is implemented inside network devices to predict flow length and improve downstream tasks performance.

**Spada** — Network monitoring ([GitHub](#))

Novel framework for network measurements. Probabilistic counting data structures (e.g., Datadog's DDSketch) are allocated on a per-flow and per-counter basis.

**SchopGPT** — Large Language Models ([GitHub](#))

Argumentative LLM agent. Arguments are refuted based on Schopenhauer's dialectic, learned via fine-tuning (HuggingFace & Mistral hackathon).

## SKILLS

### Programming

- Python (6 years)
- SQL (3 years)
- PySpark (2 years)

### Data

- PyTorch, transformers
- Sk-learn, Pandas, Numpy
- Hive, MongoDB, Databricks

### Infrastructure

- Databricks
- Azure Cloud
- AWS SageMaker

### Certifications

- Azure Fundamentals
- HubSpot and Marketo
- Google Analytics
- Excel

## LANGUAGES

- French: native
- English: advanced
- Spanish: intermediate
- German: beginner

## ACTIVITIES

- Bouldering (5 years)
- Boxing (3 years)
- Swimming (lifeguard qual.)
- Summer camp counselor

## OPEN-SOURCE

- Wikipedia contributor
- *River* stream learning library documentation

References available upon request.