Camille Barboule

5 rue Carvès, 92120 Montrouge (France) • +33(0)6 81 70 39 10 • camille.barboule@gmail.com

EDUCATION

2020-2022 **IMT ATLANTIQUE, ENGINEERING DIPLOMA** **Brest, France**
Machine Learning, Deep Learning, Advanced Mathematics, Optimized Deep Learning (MicroNet challenge: regularization, pruning and quantization techniques to reduce the floating point number and number of parameters), Design of Communicating Objects (Transcripen project)

Apr-July 19 *(1 semester)* **OTTO-FRIEDRICH-UNIVERSITÄT BAMBERG, ERASMUS PROGRAM** **Bamberg, Germany**
Finance, Economics. Thesis on "Income distribution across social strata: understanding the limits and opportunities of wealth mobility through a comparison of the Boltzmann-Gibbs and Pareto Models"

Sept-Dec 18 *(1 semester)* **UNIVERSITY OF CAMBRIDGE, EXCHANGE PROGRAM (HOMERTON COLLEGE)** **Cambridge, UK**
Finance, Economics

2017-2021 **GRENOBLE ECOLE DE MANAGEMENT, GRANDE ECOLE DIPLOMA** **Grenoble, France**
Thesis on "Improving Stock Return prediction's traditional Algorithm of Fama & French thanks to recurrent neural networks"

2015-2017 **PIERRE DE FERMAT, CLASSE PREPARATOIRE ECS** **Toulouse, France**

2015 **BACCALAURÉAT, SCIENTIFIC SECTION, OBTAINED WITH HIGH HONORS** **Toulouse, France**

EXPERIENCE

Sept22-now *(CDI)* **ORANGE** **Chatillon, France**

Machine Learning Researcher – collaboration with ISIR (former LIP6)

- SOTA on Visually-Rich Document Understanding methods: https://arxiv.org/pdf/2501.02235 comprehensive overview of state-of-the-art approaches, emphasizing their strengths and limitations and proposing promising research directions.
- Research project on handling documents with LLMs: adding the 2D position of tokens (bounding boxes from PDFs, PPTs, or OCRized documents) in the attention scores of transformers with RoPE, to better understand tables and charts within documents
- Research project on position encoding: SOTA on position encoding techniques in transformers (absolute & relative), explored techniques to mitigate positional bias (lost-in-the-middle effect, attention sink effect) in transformers, and explored techniques to isolate position information within transformers to make this position explicit and manipulable
- Research project on explainable NLP: worked on mechanistic interpretability, exploring transformers latent space using dimensionality reduction techniques (PCA) to identify concept encoding by the network on intermediate layers

Data Scientist

- Research Project about LLM adaptation to the telecommunication industry which aims at getting a "small" (7b-parameters) model performing as well as big LLMs on telecom use-cases:
  - ✓ Explored best practices about fine-tuning for domain adaptation of LLMs: on the data, preprocessing, continual pretraining (self-supervised fine-tuning on raw domain texts), instruction-tuning (on telco instructions (QA, MCQs, Summarization)), fine-tuning process itself (packing vs non-packing, loss computed on all tokens (auto-regressive way) vs on output tokens only, ...), parameter-efficient fine-tuning methods (lora vs qlora vs full fine-tuning)
  - ✓ Implemented a pipeline for domain-adaptation of LLM (from data collection, preprocessing, to fine-tuning, to results evaluation)
  - ✓ Trained models on a SLURM internal cluster using several parallelization methods
  - ✓ Wrote research paper about this work: https://arxiv.org/abs/2412.15891

Apr21-Aug22 *(Apprenticeship)* **DELPHA.IO** **Paris, France**

Machine Learning Engineer

- Developed an algorithm to perform large-scale string comparison via an optimisation of sparse matrix multiplication (vectorization of names + cosine similarity computation) to compare clients names from Salesforce environments databases in order to detect duplicates client names into their database as well as linking client names from the database which have a specific relationship (parent/subsidiary relationship, same group, …)
- Deployment of algorithms developed on AWS: use of Appflows to get/send back the data (storage in S3), AWS EMR to make an initialisation process running, AWS lambda for batch processes, API endpoints construction for Salesforce, creation of AWS stacks on CloudFormation
- Product owner: worked on an Agile way (SCRUM) to define each task / the timeline / the difficulties on both projects described above

Jan 20-Jun 20 *(6 months)* **ING BANK** **Paris, France**

Leveraged finance intern

- Financial analysis (P&L, Cashflow Statement, Budget, deleveraging profile, current trading and exit considerations analyses)
- Built LBO models challenging the Business Plan given by the management/sponsor
- Took part in the Credit Modification process for 'Prêts Garantis par l'Etat' (PGE) during the Covid-19 crisis

July 19-Dec19 *(6 months)* **BNP PARIBAS** **Paris, France**

Leveraged Finance Origination intern

- Pitched funds on potential LBOs (new LBO transaction, refinancing, extension, restructuring)
- Financial analysis of LBO targets (comps, industry research, identified risks and mitigants, analyzed financials) and built LBO models

Dec 18-mar 19 *(4 months)* **AQUA ASSET MANAGEMENT** **Paris, France**

Asset Management internship

- Financial analysis using Factset

LANGUAGE & COMPUTER SKILLS

- French (native), English (C2), German (B1)
- Python (advanced), Pytorch (advanced), Tensorflow (intermediate), Java (intermediate), C (basis), SQL (intermediate), Linux (intermediate), AWS (basis), GCP (basis)

ACTIVITIES & INTERESTS

Feb 25 **LAURÉAT OF THE HACKATHON GENAI FOR PUBLIC GOOD** **Paris, France**

- Organised by the DINUM, this hackathon aimed at using GenAI for Document Understanding on a public use-cases (criminal record understanding to help magistrate in handling their cases). Documents were scanned and contained images. The first task was about QA on documents, so we used Colpali as a retriever to get the relevant page(s) regarding the prompt. The second task was about summarizing the document, so we finetuned GOT to convert the document into structured text then passed it to the LLM Qwen with 1M context window.

Jan 25 **FRUGAL AI HACKATHON** **Paris, France**

- Participated in the text challenge about climate quote classification. I finetuned ModernBERT on this dataset.

Mar 22 **HACKATLANTIQUE (CTF ORGANIZED BY IMT ATLANTIQUE)** **Rennes, France**

- Attend the hacking championship organized by my school (specific piece of text to find hidden on the server / behind webpages, …)

2017-2018 **COMU – DANCER AT THE COMU MUSICAL (ROCK, LINDY HOP, SALSA)** **Grenoble, France**