Maxence Debes

Data Scientist · ML Engineer

maxence.debes@polytechnique.edu | +33 7 81 32 79 22 | Paris, France
maxencedbe.github.io | maxence-debes | maxencedbe | maxencedbe

Specialized in Machine Learning and AI, with a strong academic background in Mathematics and Computer Science. I enjoy solving complex technical problems by combining mathematical rigor with advanced Python to build robust, data-driven solutions.
Currently seeking a full-time position.

Education

École Polytechnique — Polytechnic Institute of Paris 2025 – 2026
MSc in Data Science, *with highest honours*
Relevant coursework: *Advanced Learning for Text and Graph Data (ALTEGRAD, joint MVA course), Deep Learning, Reinforcement Learning, Representation Learning (MVA), Generative Models, Optimization for Data Science.*

Télécom SudParis — Polytechnic Institute of Paris 2023 – 2026
Master's degree in Engineering
Relevant coursework: *Statistics and Data Analysis, Optimization, Machine Learning and Data Mining, Stochastic Processes, Continuous Optimization for Learning and Machine Learning.*

Projects

Hi! PARIS Hi!ckathon 2025 (2nd place)
– Ranked 2nd out of 80+ teams in the 2025 edition of the Hi! PARIS Data Science hackathon. Task: Predicting student PISA scores from complex socio-economic backgrounds.
– Built a gated XGBoost model: one regressor estimates the probability of a non-zero score, a second one, trained on positive scores only, predicts its value; their product reaches an R^2 of 0.79 on the test set.

Molecular Graph Captioning (Kaggle competition)
– Built a retrieval-based architecture using contrastive learning to align molecular graph embeddings with natural language descriptions in a joint latent space.
– Implemented a dedicated graph neural network encoder with PyTorch Geometric, trained with a contrastive objective (triplet loss with hard negatives) and evaluated with BLEU-4 and BERTScore.

Streaming Safety Classifier for LLMs — Capstone Project with DragonLLM (now OVHai LLM)
– Reproduced the Qwen3Guard toxicity classifier: built a dataset of 24K toxic and safe Q&A samples, synthesized and token-level labeled with Qwen3-30B.
– Trained a lightweight per-token safety head on a frozen Qwen backbone to interrupt harmful generations in real time.

Post-Training Pipeline for a Small LLM (SFT → DPO → GRPO)
– Post-trained Qwen3.5-0.8B-Base end to end with TRL in full fine-tuning: SFT on 100K conversations, DPO on 54K preference pairs, then GRPO with a verifiable reward on GSM8K, on cloud GPUs (Modal).
– Evaluated every stage with a fixed lm-eval-harness/vLLM protocol (GSM8K, IFEval, MMLU): SFT added 15 pts on IFEval; GRPO raised GSM8K accuracy from 50.1% to 62.7% (3 seeds), confirmed by a format-only reward control.

Work Experience

Data Science Intern — ChemAI | Munich, Germany Apr 2026 – Oct 2026
– Designed the architecture of, and built, a graph database of over 30M chemical reactions.
– Ran graph-native analyses to surface insights on the data's diversity, structure and reaction patterns.
– Trained ML models that leverage the graph structure for predictive analytics on chemical reactions.

Data Governance Intern — Batigère Maison Familiale | Metz, France Jul 2025 – Aug 2025
– Migrated the company to SharePoint — adopted by 100% of employees — and designed a GDPR compliance policy reinforcing data security.

Software Engineer Intern — Atos | Metz, France Jul 2024 – Aug 2024
– Designed and implemented a telework tracking application using the JHipster application generator to facilitate accessibility to workstations.

Skills & Interests

Technical Skills: Python, Java, SQL | PyTorch, PyG, Transformers, TRL, PEFT (LoRA/QLoRA), Scikit-learn, NumPy, SciPy, Pandas, Matplotlib | vLLM, Ollama, Unsloth, Neo4j, Weights & Biases | Git, Docker, Bash, Linux, LaTeX
Soft Skills: Leadership, teamwork, adaptability, technical communication
Languages: French (native), English (C1 – TOEIC 965/990), German (A2)
Engagement & Interests: Former president of a student association, competitive handball, music