

Mike Grochmal

<mike@grochmal.org>

110 Waterside, Chesham, HP5 1PE
Buckinghamshire, United Kingdom

- Computational Physicist, Lecturer, ML Eng, semi-professional Dancer.
- Passionate to move proverbial mountains of data for the purpose of science.
- No, really, been moving out-of-core data for model training for 10+ years now.

Experience

Founding ML Engineer – ThemaAI

Since Apr 2024

dask, LLM, BERT, pytorch, aws, polars, terraform

- Designed from scratch, built, maintained, tested, and scaled the ML pipeline.
- Balanced a distributed system against LLM APIs that have overall TPM and RPM limits.
- Built from scratch the horizontal scaling architecture for the company, processing 100s of GBs of data.
- The only ML engineer in a small startup.

Visiting Lecturer – City, University of London

Since Jan 2018

jupyter, numpy, pandas, sklearn, spacy, transformers

- Lead and maintain the short course on Data Analytics and Machine Learning.
- Teach full ANNs, hyperparameter tuning, bias-variance trade-offs, and dataset size considerations.
- Built the entire material for the course over the years, material also available as an online book.

Senior Data/ML/Software Engineer – BenevolentAI

Jan 2022 → Apr 2024

pyspark, jupyter, kubernetes, pytorch, neo4j, spacy, NER, LLM

- Rebuilt the scientific literature retrieval pipeline as a distributed, eventual consistency system.
- Lead a team on the migration of data pipelines feeding the NLP models.
- Speed up RAG retrieval over 100s of GBs of corpus by distributing it over a cluster of machines.
- Scaled SVO retrieval by map-reducing graph edges instead of computing the knowledge graph by joins.

Assistant Vice President – Bank of America Merrill Lynch

Jul 2017 → Sep 2019

python, nosql, distributed-computing, microservices

- Built ETL system for exchange of risk data between old and new bank systems.
- Re-engineered data processing from an ACID based database into NoSQL.
- Designed barriers into the distributed system to ensure "no-update" consistency.

Carer (career break)

2014 → 2017

Data Warehouse Architect (contract) – The MET Group

Feb 2014 → Jul 2014

postgresql, linux, amazon-web-services, redhat, django, python, bash, git, java

- Built from scratch the infrastructure and monitoring of the warehouse.
- A quick project for a small data warehouse, fully built by four engineers.

C Developer and System Administrator – FIS/Metavante
aix, hp-ux, redhat, c, perforce, perl, oracle

Apr 2010 → Jun 2013

- Maintained the team's infrastructure, the source control system, and testing environments.
- Developed database migration scripts to install our improved software with minimal RDBMS downtime.
- Trained new teams once the company was acquired into a larger organization.
- Installed our software on customer sites, physically present at the customer site.

Education

PhD, Connectionist Systems and Chaos – Birkbeck, UoL
pytorch, latex, numpy, huggingface

Since 2017 (on hold)

MSc in Intelligent Technologies (distinction) – Birkbeck, UoL
python, oracle, sql, r, libsvm, numpy, information-retrieval, nosql, haskell

2014

Faculty of Computational Physics – Jagiellonian University
unix, linux, multithreading, pthreads

2009

Faculty of Physics – Federal University of Paraná
unix, latex, c, matlab

2007

Skills

General: Training models on differently shuffled data, with observability and dynamic infrastructure.

Data: Out-of-core, (No)SQL/Graph; several times made eventual consistency DBs out of filesystems.

Likes to work with: dask pyspark numpy pandas pytorch nlp transformers parquet python
kubernetes aws neo4j redis qdrant mlflow postgres terraform unix rust c vim

Notable Projects

Learn You some ML for Great Good

<https://learnyousomeml.com/>

Online book based on student feedback on the lectures I deliver.

Style Based Artwork Classification

<https://github.com/grochmal/capybara>

Use of complexity features to classify art style, notably Kolmogorov complexity estimates.

lemannl

<https://gitlab.com/grochmal/lemannl>

Educational ANN library, with its own autograd implementation, I oft use for PyConUK talks.

Writing/Publications

q-entropic Forms

<https://grochmal.org/math/q-probability.html>

Non-ergodic entropy of a predictive system.

Strange Attractors in Stocks

<https://grochmal.org/math/sa.html>

It is surprisingly easy to find Chaos in FTSE250.