Hakob Petrosyan

ML/DL Engineer

+374 (99) 011-971
jacpetrosyan@gmail.com

Portfolio
github.com/hakob-petro
linkedin.com/in/jacpetro

Researcher with a demonstrated history of working in the computer software industry. Strong research professional with a Master's degree focused on applied mathematics and computer science from the Moscow Institute of Physics and Technology. The winner of the final part of the Republican Olympiad of Armenia in the disciplines of Mathematics, Physics, and Astronomy.

Skills

FRAMEWORKS	Python, C++, PyTorch, XGBoost, Scikit-Learn, Numpy, Pandas, Matplotlib, OpenCV, OpenGL, Python asyncio, Python Threading, HF PEFT, HF TRL, LangChain, Gym/Gym-Trading, ccxt, Backtrader
TOOLS	Git, LaTeX, Linux, Docker, TensorFlow-serving, vLLM, TEI, Tailscale/Headscale, OpenMP, MPI, DeepSpeed/FSDP, HF Accelerate
FUNDAMENTAL	Calculus, Linear Algebra, Math Optimization, Probability Theory, Theory Of Information, Network Technologies [1-4 lvl. OSI], Distributed Training/Inference of LLMs, Finances/Economics
LANGUAGES	English (upper-intermediate), Russian (fluent speaker), Armenian (native language)

Technical Experience

ML/DL Lecturer — **Sep 2024 — Present**
Yerevan State University — *Yerevan, Armenia*

> Voluntarily teaching ML/DL fundamentals to students in technical fields, aiming to advance AI education.

PyTorch Developer — **Aug 2023 — Present**
ScaleGenAI — *Palo Alto, CA, USA*

> Parallelized LLM training using Tensor, Pipeline, and Data parallelism techniques with PyTorch Pippy, Megatron-LM, DeepSpeed, and FSDP, optimizing with gradient checkpointing, mixed precision, and gradients/weights sharding and others.
> Benchmarked ASR preprocessing speed for custom and Nvidia Nemo Conformer implementations, improving training time 3x with Scaletorch's DAPP (DLOP) technology.
> Fine-tuned (with SFT, DPO, ORPO) of popular open-source LLMs (Llama/Mistral family models and etc.) using HF PEFT techniques such as LoRA/qLoRA and their variations, IA3, Llama-adapters, etc. Speeding up training with Flash-Attention 2, Unsloth, Liger kernels.
> Contributed to the full development cycle of a cost-efficient, multi-cloud LLM training/inference product. Built throughput-optimized clusters, developed an OpenAI-compatible FastAPI platform, and set up deployment with vLLM/TEI. Designed cluster fetching/provisioning logic and a backend framework for LLM fine-tuning.

Data Scientist — **Feb 2022 — May 2023**
Intelligent Solutions — *Moscow(Skolkovo), Russia*

> Color autocorrection of images using the X-Rite color palettes with OpenCV and Python colour module.
> Image white balance correction with U-Net.
> Writing an HTTP server, which was a linked server between the backend server and the TensorFlow-serving service.
> Conversion of images of dry surfaces to wet using generative models, as well as using 3D modeling methods.
> Classification of lithotypes using XGBoost and label propagation with self-learning. Comparison of supervised and semi-supervised approaches.

ML Researcher (intern) — **Oct 2020 — Jul 2022**
Acronis — *Moscow, Russia*

> Determination of the causes of the most common hardware breakdowns of systems with Windows-based operating systems. Modeling the future behavior of selected characteristics of the system.

Education

Master, *Moscow Institute of Physics and Technology* — **Aug 2022 — July 2024**

◇ Faculty of Innovation and High Technology.
◇ Department of Digital Transformation Technologies.
◇ Research topic: *"RL based dynamic determination of stop-loss limits for orders"*.

Bachelor, *Moscow Institute of Physics and Technology* — **Aug 2018 — July 2022**

◇ Faculty of Radio Engineering and Cybernetics.
◇ Department of Theoretical and Applied Informatics (Acronis).
◇ Research topic: *"Anomalies detection in the working of hardware components of PC"*.

Hakob Petrosyan

ML/DL Engineer

+374 (99) 011-971
jacpetrosyan@gmail.com
Portfolio
github.com/hakob-petro
linkedin.com/in/jacpetro

Projects

Several pet-projects that were carried out independently for interest or at the institute:

> *An implementation of the CNN SincNet architecture using PyTorch.*
> *A simple flappy bird game with NeuroEvolution of Augmenting Topologies.*
> *Prediction of missing DNA nucleotides using bidirectional LSTM.*
> *NBA players tracking using YOLOv8.*
> *Writing custom PyTorch optimizer for projected gradient descent algorithm.*
> *Realization of new NN training algorithm forward-forward using PyTorch.*
> *A cache-friendly block matrix multiplication.*
> *The numerical solution of the transport equation using MPI.*
> *The numerical integration using pthreads on shared memory.*
> *Forecasting the price of stock trades with a horizon of 1 second using XGBoost based on high-frequency data from Reuters.*

Certifications and Courses

- Neural networks and computer vision (Samsung Research, Russia)
- CCNAv7: Introduction to Networks (Cisco) :
- CCNAv7: Switching, Routing, and Wireless Essentials (Cisco)
- C++ Development Basics: White Belt (MIPT & Yandex)
- C++ Development Basics: Yellow Belt (MIPT & Yandex)
- C++ Development Basics: Red Belt (MIPT & Yandex)