

# Lavrentii Grigorian

ML/DL Engineer

Born on: 01/06/2000

Email: glavrentiy123@gmail.com

Mobile: +49 157 56537372

LinkedIn

GitHub

## INTRO

Self-motivated and enthusiastic individual seeking for a challenging **part-time** position of **ML/DL engineer**. Skilled in ML, NLP, CV and MLOps with nearly **4 years of working experience**.

## CORE COMPETENCIES

Natural Language Processing ◦ Deep Learning ◦ Machine Learning ◦ Large Language Models ◦ OOP  
A/B testing ◦ Data Preprocessing ◦ Data Analysis ◦ Computer Vision ◦ Linux ◦ CI/CD

## TOOLS

Python ◦ Git ◦ PyTorch ◦ TensorFlow ◦ PyTorch Lightning ◦ Transformers ◦ Unsloth ◦ Axolotl ◦ TGI ◦ vLLM  
◦ Pandas ◦ NumPy ◦ SciPy ◦ Docker ◦ Docker Compose ◦ Kubernetes ◦ SQL ◦ Linux ◦ MPI ◦ CUDA ◦ Label  
Studio ◦ Grafana ◦ Prometheus ◦ SuperSet ◦ PostgreSQL

## SOFT SKILLS

- Team work
- Ability to work remotely
- Critical thinking
- Attention to detail
- Project management
- Mentoring
- Continuous learning

## EDUCATION

- **University of Hildesheim** Hildesheim, DE  
*MSc, Data Analytics* Apr 2024 - Present
- **Financial University** Moscow, RU  
*BSc, Economics* Sep 2017 – July 2021

## WORK EXPERIENCE

- **Neiro.ai** Yerevan, AM  
*ML/NLP Engineer (Full-time | Remote)* Sep 2023 – Present
  - Collected, cleaned, labeled, and analyzed open-source and in-house data for LLM fine-tuning for AI Character role-play mobile application
  - Implemented LLM inference using modern backends such as TGI and vLLM; contributed a new LLM generation parameter to the TGI repository
  - Developed and continuously upgraded LLM training pipelines to accommodate growing data sizes
  - Designed and conducted A/B tests for all developed features
  - Developed a variety of NLP applications such as retrieval systems and classification
  - Administered development and production Linux machines with 4 and 8 GPUs, respectively
  - Developed and maintained CI/CD pipelines for various applications
  - Reviewed and presented new Deep Learning (LLM) research papers, enhancing team expertise and integrating the latest research into projects
  - Collaborated with NLP engineers, backend, and mobile developers to integrate complex features, enhancing the user experience of the application
  - Consistently conducted code reviews for co-workers, ensuring adherence to code style standards, identifying and fixing bugs, and enhancing overall team expertise

- Bank GPB** Moscow/Yerevan, RU/AM  
*Data Scientist/NLP Engineer (Full-time | Remote)* *June 2021 – Sep 2023*
  - Analyzed user behavior in a mobile banking application and conducted deep research, leading to crucial insights that enhanced app usability and user engagement
  - Applied graph ML methods to train client embeddings as additional features for multiple ML models
  - Developed diverse ML models for business process optimization, including employee outflow prediction, document classification, document OCR, news articles classification, NER for financial documents and machine translation, significantly improving operational efficiency and data processing accuracy and reducing overall cost
  - Collaborated with a team of DevOps professionals to deploy and maintain advanced ML models, ensuring robust performance in production environments, and shared insights at internal meet-ups and conferences, promoting a culture of continuous learning and innovation
  - Optimized DL model inference speed using pruning, knowledge distillation and ONNX
- Freight One** Moscow, RU  
*Data Scientist, University Internship (Part-time | On-site)* *Jan 2021 – June 2021*
  - Contributed to the development of railway wagon downtime prediction model via data analysis, feature engineering and statistical hypothesis testing
  - Reported the results to colleagues and stakeholders

## ADDITIONAL EDUCATION

---

**Sep 2023**      **CS224N: Natural Language Processing with Deep Learning.** Stanford Online [\[Github\]](#)  
**Dec 2022**      **Machine Learning Specialization.** DeepLearning.ai & Stanford [\[Certificate\]](#)

## CONFERENCES & COMPETITIONS

---

**June 2023**      **Generalization Challenge.** Hackathon, Sibur [\[Github\]](#)  
**June 2023**      **Text Deduplication: Review.** Conference, Data Fest ODS [\[Program, Presentation\]](#)  
**March 2023**    **Testing Machine Learning Systems.** Conference, OpenTalks.ai [\[Program, Presentation\]](#)

## LANGUAGES

---

- English – C1 (IELTS 7.5)
- German – B1.1 (Learning)
- Armenian – native
- Russian – native

## REFERENCES

---

- Oleg Khomenko, CEO – Neiro.ai [\[Letter\]](#)
- Vladimir Dashkovsky, Head of ML – Bank GPB [\[Letter\]](#)