ALEXEY PAVLOV

tg: @glaucomaa | github: glaucomaa | aamocualg@gmail.com

WORK EXPERIENCE

Data Scientist/Engineer 2024 - present
Best Place

- Implemented ML service for product categorization (*Sentence-BERT*): F1 ↑ from 74% to 88%, processing time ↓ by 5x.
- Automated categorization pipeline (*RuBERT*), reducing manual labeling by 80%.
- Trained demand forecasting model (*Prophet* + *CatBoost*): MAPE ↓ from 18% to 9%.

ML / NLP Research Engineer 2025 – present
EMS VK (big tech company)

- Developed open framework for training and experiments: modular refactoring, logging support, reduced disk space usage by 60%.
- Formulated and prioritized hypotheses (single/multi-head attention, Local–Global schemes, RoPE variants); prepared test plans together with PI.
- Created and maintained experiment configurations (Hydra / YAML / JSON / Bash); automated massive ablation studies on GLUE and LRA - LRA accuracy ↑ 1.5 pp.

PROJECTS

Streaming ETL Platform for Entity Management

- Designed containerized architecture (PostgreSQL + 3-broker Kafka cluster + Cassandra + Spark) with Docker Compose and persistent volumes.
- Implemented REST API on FastAPI with async connection pooling; increased throughput to 3,000 requests per second.
- Built reliable ETL pipeline: atomic JSON format saving, data streaming through Kafka and writing to Cassandra.

RAG vs LoRA

- Developed RAG system based on *Falcon-1B* with dynamic context search, improving average text generation quality by +**27%** across 8 metrics (BLEU, ROUGE-1/2/L, METEOR, Precision/Recall/F1).
- Reduced VRAM requirements by **3x** through QLoRA while maintaining >**90%** quality.
- Created benchmark on **12,000** QA examples from 3 datasets, confirming RAG approach effectiveness for question-answering systems.
- Proposed and compared **two segmentation algorithms**, showing superiority of the first over static approaches by 2 pp and lagging behind ideal approach by only 7 pp.

EDUCATION

Higher School of Economics, Saint Petersburg (Master's degree),
School of Physics, Mathematics and Computer Science,
Machine Learning and Data Analysis 2024 - 2026
GPA: 8.5/10.0

Saint Petersburg State University, Saint Petersburg (Bachelor's degree),
Faculty of Applied Mathematics and Control Processes, FIIT,
Programming and Information Technologies 2020 - 2024
GPA: 4.76/5.00

PUBLICATIONS

- Methods of Fine-Tuning: RAG vs LoRA - CPS'24 conference, HIT, Harbin

SKILLS

Python, C/C++, Rust, PyTorch, Transformers, scikit-learn, CatBoost, ONNX Runtime,
Docker, FastAPI, Kafka, ClearML, Hydra, Linux, Git, SQL, LoRA, RAG, Quantization, Pruning