

EDUCATION

Bachelor in Computer Science, HSE University 2023-2027

*Applied Mathematics and Informatics, **Budget**, GPA 7/10*

Main courses: Linear Algebra, Calculus, Python programming, Computer architecture, ML/DL, Data Structures, Data Analysis etc.

Efficient DL Systems 2023-2027

YSDA and HSE course

Worked with ONNX, DVC, model quantization, experiment tracking via MLFlow, wandb and dvc, made efficient deploy GPT2 with Docker, Grafana. Finetuned 3B model via Model-parallel. Optimized inference with OpenVino.

DL Courses 2022

- **DLS** : Introduction to DL course. Successfully completed the tasks based on CNN's: Image classification, Image segmentation, Image classification
- **Tinkoff DL/DL Advanced** ([link](#)): Successfully completed labs and mini-projects with Neural Networks: Transformer models, Metric Learning, Language Modelling, RNN, AE/VAE, GAN, Segmentation, Translation, Effective Learning and Deploy
- All my 10+ SQL, AI, Programming competitions and courses diplomas: ->[link](#) <-

EXPERIENCES

BigData Analyst, Beeline Kazakhstan 2023

Beeline is the largest Fintech company in Kazakhstan

- Interacted with production DB's based on HDFS, PySpark, Kafka, PostgreSQL to upload and send data.
- Developed and brought to production SQL triggers for MyBeeline App.
- Refactored Multisim Model, increased performance for 30 percents(F1 score)
- Communicated with customers, promptly and successfully solved existing and new issues.

DANO Olympiad Winner, HSE and Tinkoff 2022

Data Analysis National Olympiad

- Solved different analytical, data analysis tasks
- Led team of 5 people, which made research on dataset of movie ticket sales in 2020
- Organized team and tracked a progress in Miro, Notion
- Made EDA, Feature engineering, Data vizualizaon using pandas/seaborn etc.

Totally - **23rd/7k+** place(individual ranking), **9th/56 place**(team tour)

PROJECTS

GPT2 Pretraining from Scratch 2024

- Implemented vanilla GPT2 on pure PyTorch.
- Optimized with flash attention, compiling, autocasting, right model initialization, mixed precision
- Used gradient accumulation for 0.5M batchsize effect. Trained on FineWebEDU 10B dataset with a DDP training on 2x RTX4090
- Achieved 32 percent HellaSwag score on my model vs 28 percent on original model

XmasHack Winner**2024**

- Took 1st/10 place in a hackathon aimed at optimizing marketing expenses of the largest CIS TV channel via TVR Index prediction on historical tabular data
- Used Optuna, Catboost, Complicated Feature Engineering with seasonality detection
- Achieved 23 percent MAPE at private test set. Closest to us solution had 25 percent MAPE