Roman Trostonetskiy

Machine Learning Engineer
trostobusiness@gmail.com | +7 915 106 64 87 | github.com/smurkiooo | Telegram @romantrost

Experience

SAPFIR Agency (Skolkovo) | Intern Jun–Oct 2026

- Developed an end-to-end system to track key indicators of AI development in Russia. Built data processing pipelines and LangGraph agent workflows, using LangSmith for tracing and debugging.

Selected Projects

RiskEmbed GitHub

- Trained a causal Transformer on 5.5 million monthly American Express statements to generate embeddings of customer histories for credit default prediction.
- Added embeddings to LightGBM alongside aggregate features. On held-out customers, Gini rose from 0.92295 to 0.92363 (+0.00068, paired bootstrap 95% CI: +0.00017 to +0.00115).

SmartPropertyScraper GitHub

- Built a pipeline for Cian property listings using MinIO and DVC. A CatBoost price model with PCA-compressed CLIP photo embeddings achieved a validation R-squared of 0.73 (1,374 listings in the dataset).
- Built semantic property search with FastAPI, multilingual-e5, ChromaDB and a local LLM, plus a Telegram bot for price estimates.

DocChat GitHub

- Built a local RAG chat app for PDFs with page-level citations and relevance filtering. Combined E5 and BM25 with reciprocal rank fusion and ran generation through LM Studio.
- Retrieved all reference passages in the top six results for each of 16 answerable questions on a synthetic test set.

Hackathons

AlfaFuture (Alfa-Bank) Oct 2026

- Co-developed a lending dashboard for entrepreneurs. Worked on a neural feature pipeline that used customer embeddings for credit scoring, and fine-tuned Qwen3.5-4B using PEFT.

Technoforge AgroTECH (EkoNiva, HSE University) 3rd place, Autumn 2025

- Tested hypotheses, engineered features and trained gradient boosting models to recommend dairy cow diets for a target milk fatty acid profile.

Education

Financial University under the Government of the Russian Federation 2024–2028
Bachelor's programme in Applied Mathematics and Informatics, specialising in Applied Machine Learning

Yandex Practicum 2026
ML Engineering course for experienced practitioners

BYTE Career Accelerator (Financial University, Gazprombank) 2026
Machine Learning and Deep Learning course, diploma with distinction

Skills

ML and Data Python, SQL, DuckDB, pandas, NumPy, scikit-learn, scikit-uplift, CatBoost, LightGBM, Optuna
Deep Learning and LLMs PyTorch, Transformers, PEFT (LoRA/QLoRA), RAG, vLLM, LangChain ecosystem
Infrastructure FastAPI, Docker, Airflow, Dask, Hadoop, MLflow, DVC, Prometheus, Grafana, Git
Languages English C1 (IELTS), German B2 (Goethe-Zertifikat)