

Konstantin Korolev

Email: korolev.konstantin.v@gmail.com LinkedIn: [korolevk](#) GitHub: [advpropsys](#) Telegram: [advprop](#)

SUMMARY

Machine Learning Research Engineer with commercial experience passionate about implementing creative, novel solutions in ML and quantitative research. Love startup culture, data science and good coffee.

EXPERIENCE

Alber Blanc - HFT Company [↗](#)

Remote / SPB / Lugano, Switzerland

- Data Engineer*

July 2024 — Present

[Python](#) [Agents](#) [Databases](#) [OSINT](#)

- Creating custom pipelines for web scraping and searching for unique events.
- Developed HR pipeline analytics; built an agent-based flow that analyzed 2k+ resumes and provided feedback.

VikhrModels [↗](#)

Remote

- Machine Learning Research Engineer*

April 2024 - Present

[Python](#) [TRL](#) [Torch](#) [Infra](#)

- Developing best-in-class large language models for specific languages.
- Conducting post-training research and inference optimization for medium-sized models.
- Created more than 50 high-quality synthetic datasets.
- Developed the most popular LLM benchmark for the Russian language, **Arena_Hard**, as part of the team.
- 100k+ downloads, 200+ HF likes, 100+ GitHub stars.

House of Pitch - Backed by Libermans Brothers, Stanford [↗](#)

Remote / Los Angeles, USA

- Machine Learning Engineer*

July 2023 — July 2024

[Python](#) [Pandas](#) [PyTorch](#) [FastAPI](#) [Databases](#) [Docker](#) [CI/CD](#) [Prefect](#)

- Executed data engineering and mining tasks—including website scraping, dataset creation, data validation, API reverse engineering, and data pre-processing to achieve speeds of up to 800 articles per minute, outperforming open-source web scrapers.
- Developed a custom gradient-free ranking algorithm for sparse user-user interactions, leading to a 60% improvement in key performance metrics.
- Tuned language model ensembles for pitch analysis.

MOEX (HSE Lab) [↗](#)

Remote / Moscow, Russia

- Quantitative Researcher Intern*

Nov 2023

[Python](#) [Pandas](#) [XGBoost](#) [LightGBM](#) [CatBoost](#)

- Performed pre- and post-processing of Crypto trade data, generated features, and built a market model.
- Tasked with processing Binance order book data and researching the Lead-Lag market model, achieving results comparable to those in this paper [↗](#).

EDUCATION

- **Higher School of Economics, FCS** [↗](#)

Course: Computer Science and Data Science

GPA: 8

PROJECTS

- **Venture Capital Research Platform (Merged with House of Pitch)**

[Torch](#) [React](#) [LLM](#) [Distributed Computing](#)

- * Developed a platform for rapid text analysis of pitches by leveraging distributed inference of custom fine-tuned LLMs.
- * Created data pipelines that parsed over 5TB of company data.

CONFERENCES & HACKS

- Second author in LLM research [paper](#) at MRL (EMNLP Workshop 2024), Miami, USA. (Cited in A*)
- International Young Scientists Conferences 2022, 2023, Serbia — gold and bronze medals in Computer Science and Engineering.
- Uber Global Data Hackathon 2023 — LATAM regional winner.
- Served as a Junior Math Expert at MSU during an international high school and student conference for 300+ students.