Aleksandr Popkov

Location: Saint Petersburg, Russia
Email: alr.popkov@gmail.com
GitHub Telegram LinkedIn

7 YoE in ML, including 2D CV and NLP. I have experience as a data scientist, an MLE with a focus on ML system design, and a TechLead in a team of up to 3-4 people. For over three years, I have combined industrial development with academic research (publications in Q1/Q2 journals), and I contribute to open source.

INTERESTS

Efficient ML, interpretable ML, ML/AI system design, NLP, multimodal learning, LLM agents, deep learning in CV, math reasoning of language models

EXPERIENCE

Research Associate   **Feb 2024 – Present**

Zoological Institute of Russian Academy of Sciences, Saint Petersburg, Russia

- Automated the process of image annotation and segmentation using LabelStudio, MinIO, and the Segment Anything Model (SAM), reducing the time it takes to prepare datasets for micrograph analysis severalfold.
- Developed a pipeline for clustering Lycaenidae wing images in the UV spectrum based on encoder-decoder architectures and BioCLIP, ensuring experimental reproducibility.

Sr. Data Scientist   **Aug 2022 – Oct 2025**

"GazpromNeft Regional Sales", LLC, Saint Petersburg, Russia

- **RAG & Prototyping**: Designed and implemented a prototype of an enterprise Q&A RAG system for the Confluence knowledge base based on **LangChain** from scratch, including evaluation process with **Langfuse** and **LabelStudio**.
- **NLP & LLM**: Automated categorization of HR interview texts (reduced processing time from 2-7 hours to $\sim$ **5** minutes) and accelerated the analysis of corporate client phone reviews by at least **10** times using topic classification (HF transformers, ONNXRuntime); conducted LLM training on a BERT-like architecture for the NER task (entity-based F1-score 0.67-0.92).
- **Business Impact**: Reduced time spent on colleagues summary meetings by more than $\sim$8 times due to the implementation of oil production forecasting by wells (scikit-learn, polars) (MAPE 1-11%, in **2/3 of cases better than experts**).
- **Efficient Engineering**: Implemented the practice of uplift modeling on decision trees, while simultaneously reducing exponentially growing computational costs to linear ones in **causalml**.
- **Pragmatic Engineering**: Implemented a simple non-ML solution (pandas, PyQT), which reduced colleagues' time spent by $\sim$**12** times (from 2 hours to 10 minutes) per survey.
- **ML Infrastructure:** Proposed a development strategy for an ML platform based on Kubeflow and reviewed support for various GPUs, which became the core of the company's infrastructure project.
- **Team Development:** Proposed and organized a working messenger for ML/DS based on Mattermost, which developed into a developer communication service

project; Organized and led meetings of the internal ML club on NLP and AI-assisted coding (5+ participants).

Research Laboratory Assistant   Jun 2020 – Dec 2022
Zoological Institute of Russian Academy of Sciences, Saint Petersburg, Russia

- Developed a service for recognizing agricultural pests *Eurygaster* spp. (F1-score 0.95+ on the museum collection dataset), deployed with BentoML.
- Organized a full cycle of experiments on classifying images of the museum's Heteroptera and Coleoptera collection using CNN (PyTorch), the results of which formed the basis for the publication.

Data Scientist   Nov 2018 – Aug-2022
"GazpromNeft Regional Sales", LLC, Saint Petersburg, Russia

- Developed an IoT computer vision system for gas stations based on NVIDIA Jetson and the Kubernetes-certified platform.
- Developed the Marketing Mix Modeling model (Prophet, Stan).
- Developed and conducted a workshop on introducing 2D CV architectures.

EDUCATION

Saint Petersburg State University, Saint Petersburg, Russia   2020
MSc in Business Informatics

Saint Petersburg State University, Saint Petersburg, Russia   2018
BSc in Economics, Mathematical and Statistical Methods

PUBLICATIONS

- Namyatova, A. A., Dzhelali P. A., Tyts, V. D., & **Popkov, A. A.** (2024). Climate change effect on the widely distributed Palearctic plant bug species (Insecta: Heteroptera: Miridae). *PeerJ 12:e18377*
- **Popkov, A.**, Konstantinov, F., Neimorovets, V., & Solodovnikov, A. (2022). Machine learning for expert-level image-based identification of very similar species in the hyperdiverse plant bug family Miridae (Hemiptera: Heteroptera). *Systematic Entomology, 47*(3), 487–503

SKILLS

Python — Cython — C — bash — Linux — PyTorch — TensorFlow — Hugging Face — LangChain — LangFuse — NLP — HF transformers — SpaCy — LLM — RAG — Computer Vision — Kubeflow — Triton Inference Server — ONNX — Docker — k3s — PostgreSQL — Vector DB (Milvus, ChromaDB) — Quantization — EfficientML — NVIDIA Jetson

LANGUAGES

- Russian (native) • English (fluent, C1) • Japanese (beginner) • German (beginner) • French (beginner)