Aleksandr Popkov

Email: alr.popkov@gmail.com

7 YoE in ML, including 2D CV and NLP. I tried the roles of a Data Scientist, a MLE with involvement in ML system design, and a Tech Lead of a team of up to 3–4 people. 3 years of experience in research projects in biology related to computer vision. ML open-source contributor.

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 Segment Anything Model (SAM), reducing the time to prepare the dataset for model training.
- Developed a pipeline for clustering images of Lycaenidae wings in the UV spectrum based on encoder–decoder architectures and BioCLIP, ensuring reproducibility of experiments and scalability of processing.

Sr. Data Scientist **Aug 2022 – Present**

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

- **LLM:** Proposed and developed the first prototype of a corporate QA RAG system for the Confluence knowledge base based on LangChain, including a quality assessment process with Langfuse and LabelStudio.
- **NLP:** Automated the categorization of HR interview texts (reduced processing time from 2-7 hours to $\sim$ 5 minutes); accelerated the analysis of corporate client phone reviews by at least 10 times using topic classification.
- **Analytics and ML:** Developed a corporate tool for A/B testing; Implemented the practice of uplift modeling on decision trees, simultaneously reducing exponentially growing computational costs to linear ones in causalml; Reduced the time spent on summary colleague meetings by more than $\sim$8 times due to the implementation of oil production forecasting by wells (MAPE 1-11%, in 2/3 of cases better than experts).
- **Like at Google Research**: Found an effective solution for validating corporate survey results without ML using pandas and PyQT, which reduced colleagues' time by a factor of 12 (from 2 hours to 10 minutes).
- **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.

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., using BentoML.

- Responsible for the ML infrastructure and the full cycle of experiments to classify images of the museum's Heteroptera and Coleoptera collection using CNN.

Research Intern **Sep 2020 – Dec 2020**
Saint Petersburg State University, Saint Petersburg, Russia

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 a forecasting system based on probabilistic programming (Prophet, Stan).
- Developed and conducted a workshop on introducing 2D CV architectures.

Analytics Intern **Feb 2017 – Apr 2017**
"GazpromNeft Center", LLC, Saint Petersburg, Russia

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
- PyTorch, TensorFlow 2
- Databases: RDB (PGSQL, MSSQL, MySQL), VDB (Milvus, ChromaDB)
- Docker, k3s, Kubeflow
- Edge-computing: NVIDIA Jetson Nano, NX, AGX Xavier

LANGUAGES

• Russian (native) • English (fluent) • Japanese (beginner) • German (beginner) • French (beginner)