

Александр Попков

Email: alr.popkov@gmail.com

Желаемая позиция: Sr./Lead DS/MLE

7 YoE в ML, включая 2D CV, NLP. Имею опыт в роли DS-специалиста, MLE с погружением в дизайн ML систем, TechLead-а в команде до 3-4 чел. 3 YoE в исследовательских проектах в биологии с приложением CV. Участвую в развитии ML open-source.

Области интересов

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

Опыт работы

Мл. научный сотрудник

Фев 2024 – н.вр.

Зоологический институт РАН, Санкт-Петербург, Россия

- Автоматизировал процесс аннотации и сегментации изображений с использованием **LabelStudio**, **MinIO** и **Segment Anything Model (SAM)**, сократив время подготовки датасета для анализа микрофотографий.
- Разработал пайплайн для кластеризации изображений крыльев *Lycaenidae* в УФ-спектре на базе encoder-decoder архитектур и **BioCLIP**, обеспечив воспроизводимость экспериментов и масштабируемость обработки.

Sr. Data Scientist

Авг 2022 – Окт 2025

Газпромнефть, Санкт-Петербург, Россия

- **LLM:** Предложил и разработал первый прототип корпоративной Q&A **RAG**-системы для базы знаний Confluence на основе **LangChain**, включая процесс оценки качества с **Langfuse** и **LabelStudio**.
- **NLP:** Автоматизировал категоризацию текстов HR-интервью (снизил время обработки с 2-7 часов до ~ 5 минут); ускорил анализ телефонных отзывов корпоративных клиентов не менее чем в **10** раз с помощью классификации тем (HF transformers, ONNXRuntime).
- **CV:** Подготовил техническое предложение по созданию системы видеоаналитики для хоккейных матчей, включая архитектуру и ключевые алгоритмы.
- **Аналитика и ML:** Разработал корпоративный инструмент для A/B-тестов; Внедрил практику uplift-моделирования на деревьях решений, параллельно снизив экспоненциально растущие затраты на вычисления до линейных в **causalml**; Сократил затраты времени на суммарные совещания коллег более чем в ~8 раз за счёт внедрения прогнозирования добычи нефти по скважинам (MAPE 1-11%, в **2/3** случаев лучше экспертов).
- **Как в Google Research:** Нашел эффективное решение для валидации результатов корпоративных опросов без ML с pandas и PyQT, что снизило время затрат коллег в ~12 раз (с 2ч. до 10мин.).
- **ML-инфраструктура:** Предложил стратегию развития ML-платформы на базе **Kubeflow** и провел ревью поддержки разных GPU, что стало основной для инфраструктурного проекта в компании.

- **Экспертиза:** Проводил разбор задач и code review для проектов Citizen Data Science (до 3 задач в месяц), повысив качество решений.
- **Развитие команд:** Предложил и организовал рабочий мессенджер для ML/DS на базе Mattermost, что развилось в проект сервиса коммуникации для разработчиков; Организовывал и проводил встречи внутреннего ML-клуба по NLP и **AI-assisted coding** (5+ участников).

Лаборант-исследователь

Июн 2020 – Дек 2022

Зоологический институт РАН, Санкт-Петербург, Россия

- Разработал сервис распознавания с/х вредителей *Eurygaster* spp. с BentoML.
- Отвечал за ML-инфраструктуру и полный цикл проведения экспериментов по классификации снимков музейной коллекции Heteroptera, Coleoptera с CNN.

Data Scientist

Ноя 2018 – Авг-2022

Газпромнефть, Санкт-Петербург, Россия

- Разработал IoT систему компьютерного зрения для заправочных станций на базе NVIDIA Jetson и **k3s Kubernetes**-certified платформе.
- Разработал систему прогнозирования на базе вероятностного программирования (Prophet, Stan).
- Разработал и провел воркшоп по введению в архитектуры 2D CV.

Образование

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия 2020

Магистр, Бизнес-информатика

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия 2018

Бакалавр, Экономика, Математические и статистические методы

Публикации

- 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

Навыки

- Python, Cython, C, bash
- PyTorch, TensorFlow 2
- HF transformers, SpaCy
- Triton Inference Server, TensorRT, ONNXRuntime, llama.cpp
- Databases: RDB (PGSQL, MSSQL, MySQL), VDB (Milvus, ChromaDB)
- Docker, k3s, Kubeflow
- Edge-computing: NVIDIA Jetson Nano, NX, AGX Xavier

Языки

English (fluent, TOEFL 106/120) / Japanese (beginner) / German (beginner) / French (beginner)