Марк Ермолаев

ML Engineer / Data Scientist (стажёр)
Долгопрудный | +7 926 494-51-50 | ermolaev.ma@phystech.edu

О себе

Студент 3 курса МФТИ, стажёр ML-команды Сбера и стажёр-исследователь Сколтеха. Задачи: детекция аномалий во временных рядах, NLP, LLM-агенты. Сильная сторона — диагностика: нахожу неочевидные ошибки в моделях и подтверждаю их на данных. Приоритеты: Classic ML, CV, NLP, DL.

Опыт работы

Стажёр ML / Data Science — Сбер Москва
ML-команда: детекция аномалий и AI-агенты *Авг 2026 — наст. время*

- **Детекция аномалий во временных рядах** (прод, команда из 2 человек; бустинг на лаговых и календарных признаках + IsolationForest). Нашёл ошибку калибровки σ: разброс оценивался на горизонте $\sim$107 шагов рекурсивного прогноза, а применялся на $\sim$7 — полосы детекции были в $\sim$4 раза шире нужного и пропускали аномалии.
- Собрал replay-стенд на исторических рядах: показал, что серия из 25 рабочих дней без поставки данных прошла незамеченной. Нашёл 4 источника утечки из будущего и неочевидные баги (`argmin` вместо `idxmin` при выборе порядка AR, парсер, молча терявший числа вида 1e-5).
- **LLM-агент проверки транзакций** (FastAPI, asyncio, LangChain, GigaChat): устранил причину 100% промахов поиска по справочнику (явный Pydantic-контракт вместо нарезки строк); разобрал 4 дефекта ретраев при HTTP 429 (общий счётчик бэкоффа между корутинами, thundering herd) и подготовил фикс; обосновал $F_{0.5}$ как базовую метрику: ручная проверка ограничена, важнее precision.
- **NLP-детектор нетипичных проводок** (novelty-скоринг токенов, миллионы проводок в сутки): согласовал с техлидом и заказчиком доработку — лемматизацию, чтобы словоформы известных слов не давали ложных срабатываний.

Стажёр-исследователь (НИР) — Сколтех Москва
Лаборатория машинного обучения для материаловедения *Авг 2026 — наст. время*

- Тема — диффузионные модели генерации кристаллов (MatterGen, MiAD; метрики S.U.N., MP-20). Воспроизвёл MatterGen: условная генерация (Ga–As), релаксация MLIP-потенциалом MatterSim, анализ симметрии и энергий; обошёл баг в модуле оценки. Написал 25-страничный гайд по HPC-кластеру (Slurm).

Стажёр-тестировщик (нагрузочное тестирование) — ПСБ Москва
Сценарии нагрузки, анализ метрик производительности, поиск узких мест *Июл 2025 — Окт 2025*

Преподаватель физики и математики Дистанционно
Частная практика: подготовка к олимпиадам и ЕГЭ *Май 2023 — наст. время*

- Ученики в 2026: четыре 100-балльника ЕГЭ и 5 дипломов олимпиад (МОШ, «Физтех», «Росатом»).

Проекты и достижения

- **Kaggle: NeurIPS — Open Polymer Prediction 2025** (курсовой проект): предсказание 5 свойств полимеров по SMILES — пайплайн из CatBoost, гауссовских процессов (GPR) и графовой нейросети в стиле CGCNN.
- Поступил в МФТИ без вступительных испытаний (БВИ) как победитель олимпиады.
- Олимпиады: «Курчатов» (физика) — победитель; «Физтех» (физика, дважды), Московская олимпиада школьников (физика), «Росатом» (физика, математика) — призёр.

Образование

МФТИ — бакалавриат Долгопрудный
ЛФИ (бывш. ФОПФ), «Прикладные математика и физика», 3 курс *Выпуск 2028*

- Матанализ, линейная алгебра, дифференциальные уравнения, ТФКП, урматы, теория вероятностей и статистика; доп. — продвинутый курс математики (НМУ), IT-курсы кафедр МОУ и АТП (алгоритмы, ML).

Школа 21 (Сбер) Москва
Data Science и DevOps: ML-проекты (feature engineering, валидация, optuna) *Июн 2025 — наст. время*

Яндекс Лицей Москва
«Промышленная разработка на языке Python» — диплом с отличием *2019 — 2021*

Навыки

Языки: Python и SQL (продвинутый), C/C++ (базовый); английский — продвинутый
Classic ML: scikit-learn, CatBoost, IsolationForest, гауссовские процессы, кластеризация (K-Means, DBSCAN), рекомендательные системы (базово); временные ряды, feature engineering, валидация, optuna
DL и LLM: PyTorch, PyTorch Geometric, диффузионные модели; LangChain, GigaChat, open-weight LLM
Разработка: FastAPI, Pydantic, asyncio; uv, ruff, pyright (strict); Git, Docker, Linux, Slurm
Данные и статистика: pandas, polars, NumPy, matplotlib; Greenplum, PySpark и Airflow (базово); A/B-тесты (t-test, bootstrap, размер выборки)