Дмитрий Иванов

ML/MLOps-инженер

Расчётное моделирование — 11 лет (с 2023 г. руковожу бюро) | коммерческий ML — 1 год (контракт, 2021–2022)
ML-системы в собственных проектах — с 2025 г.

Ростов-на-Дону, Россия | удалённо | ИП / самозанятость
Полная или частичная занятость | выход через 2 недели

damiyan.ivanov@gmail.com | +375 (29) 155-69-01
portfolio.dmitryivanov.dev/ru | linkedin.com/in/dmitry-ivanov-ds | gitlab.com/dmitry-ivanov-ds

О себе

Строю ML-системы от сбора данных и обучения моделей до сервиса в Kubernetes. Коммерческий ML — год по контракту: прототип извлечения данных из документов для Chemwatch (Австралия). С 2025 г. развиваю собственные ML-проекты, включая SoccerPredictAI. 11 лет строю и проверяю расчётные модели сельхозтехники на полевых данных: модель принимают, только когда она совпала с испытаниями. С 2023 г. руковожу бюро имитационного моделирования в Ростсельмаше. Ищу работу ML/MLOps-инженером, где эта дисциплина проверки моделей на данных нужна.

Ключевые результаты

- **SoccerPredictAI:** ML-система полного цикла, собрано 950 000+ матчей — от скрейпинга до сервиса в Kubernetes. На отложенной выборке (135 970 матчей) log-loss 1,006: лучше частотного прогноза (1,071), но хуже букмекера (∼0,97).
- **Chemwatch:** прототип извлечения состава из паспортов безопасности. На ∼1000 документах 95% извлечены верно — все компоненты, без лишних (сверка вручную); остальные 5% алгоритм сам передал человеку.
- **Валидация моделей на полевых данных:** расчёт системы очистки комбайна H820 совпал с агротехнической оценкой в поле: среднее относительное расхождение менее 5% по зависимости от подачи (10–12 режимов). В сезоне 2026 за 1,5 месяца моделирование нашло причину повышенных потерь; после доработки потери снижены в 4 раза.

Собственные ML-проекты (личные)

SoccerPredictAI — прогноз исходов футбольных матчей v1: май 2025 – май 2026; v2 — в работе
Демо | Документация (англ.)

- Данные: скрейпинг 950 000+ матчей с одного сайта статистики; коэффициенты букмекеров из 2 источников — только для сравнения с рынком, не как признаки модели. Хранение — PostgreSQL и MinIO, SQL-запросы (JOIN, агрегации) писал сам.
- Пайплайн: проверки качества данных (Great Expectations) → признаки (скользящие статистики, рейтинги Elo) только по матчам до даты прогноза → walk-forward кросс-валидация, выбор модели только по ней, отложенная выборка оценивается один раз → XGBoost, Optuna, реестр MLflow → FastAPI в Kubernetes. DVC — 22 стадии, 564 теста.
- Результат v1 на отложенной выборке (135 970 матчей с 2024 г.): log-loss 1,006 (частотный прогноз 1,071, равновероятный 1,099, букмекер ∼0,97); калибровка ECE 0,004. Симуляция ставок — −0,7%. Сервис: p50 ∼126 мс (локальный тест, WSL/Docker), инфраструктура дешевле €30 в месяц.
- v2 (с сент. 2026): переписываю с агентным процессом разработки — система разрослась, и одному стало трудно удерживать связь обучения, мониторинга и качества модели. Мониторинг дрейфа (Evidently) — в работе.

CAE Copilot — ассистент по документации CAE (MVP) июнь 2026 – н. в.
Демо | Документация (англ.)

- RAG по 9 руководствам Ansys Rocky со ссылками на страницы источника; расчётные инструменты; LangGraph, Qdrant, FastAPI, Kubernetes.
- Автоматические проверки маршрутизации, ссылок и инструментов на 14 авторских сценариях, без LLM; качество ответов не измерено. Пользователей пока нет.

Агентный процесс разработки и self-hosted платформа сент. 2026 – н. в.

- Я пишу спецификацию задачи и принимаю результат; код пишет агент, ревью спецификации и кода — модель другого вендора. Пример: ревью до кода нашло ошибку в спецификации — `git push` по тайм-ауту считался неудачей, хотя сервер мог уже принять обновление.
- Одноузловой Kubernetes, поднятый вместе с агентом: Prometheus, Grafana, Loki, алерты; ночные дампы БД; пробное восстановление выполнялось.

Опыт работы

Начальник бюро имитационного моделирования – *Ростсельмаш (Ростов-на-Дону)* февр. 2023 – н. в.
Ведущий инженер-конструктор дек. 2021 – февр. 2023

- Бюро ведёт CFD-, DEM- и RBD-расчёты комбайнов и адаптеров, ∼15–20 задач в месяц; расчёт бюро — обязательный этап согласования конструкции (APQP). Вырастил бюро с одного сотрудника до 10 инженеров; сопровождение — с 2 до ∼5 образцов техники.
- Ввёл ревью каждой расчётной модели: перед отправкой конструкторам результат сверяется с упрощённой инженерной оценкой и данными испытаний.
- Данные испытаний: агротехнические оценки, тензометрия, полевые эксперименты; обработка — Python, MATLAB, Excel. Пример: выгрузка зерна — +50% производительности минимальными доработками по расчёту и тензометрии.
- Автоматизировал ∼5 типовых операций подготовки модели скриптами на Python через API.

Data Scientist – *Chemwatch, Австралия (контракт, удалённо)* янв. 2021 – янв. 2022

- Задача: показать менеджменту, что ручное извлечение состава из паспортов безопасности (сканы и PDF без структуры) можно автоматизировать. Работал один, параллельно с основной работой; куратор заказчика каждую неделю проверял ход работы и результаты.
- Метод: OCR (AWS Textract) → разметка страницы кластеризацией (заголовки, таблицы, абзацы) → поиск нужного раздела (кластеризация + классификация) → извлечение и проверка номеров CAS по открытому реестру.
- Результат: на ∼1000 документах 95% извлечены верно — найдены все компоненты и нет лишних (сверка вручную); остальные 5% алгоритм сам отметил как нераспознанные и передал человеку. Только английский (более 80% массива); перенос на другие языки не проверялся. Прототип передан заказчику.

Ведущий инженер-конструктор – *Гомсельмаш (Гомель, Беларусь)* авг. 2012 – дек. 2021

- По своей инициативе внедрил DEM-моделирование (2015) и добился закупки расчётного комплекса — ПО Ansys и серверы (2020): написал и защитил проект.
- Спроектировал стенд для измерения свойств технологической массы: параметры DEM-модели — из измерений, а не из справочника (CADFEM Review, 2020). Обучил расчётам ∼5 инженеров.

Навыки (применённые в проектах выше)

ML: Python, pandas, NumPy, scikit-learn, XGBoost, Optuna; временная кросс-валидация, калибровка; кластеризация и классификация документов

MLOps: MLflow, DVC, Great Expectations, Airflow, Evidently (мониторинг дрейфа — в работе)

Данные, сервисы моделей и инфраструктура: SQL (PostgreSQL), MinIO, FastAPI, Celery, RabbitMQ, Docker, Kubernetes, Helm, GitLab CI; Prometheus, Grafana, Loki

LLM: LangGraph, RAG, Qdrant; разработка с агентами (OpenAI Codex, Claude)

Моделирование и данные испытаний: DEM, CFD, RBD; Ansys Rocky, Fluent, Motion, Workbench; MATLAB; тензометрия

Образование, курсы и языки

Магистратура (2016–2018) и **высшее образование** (2007–2012), машиностроение — ГГТУ имени П. О. Сухого, Гомель

Курсы: ODS.ai — LLM (2026), MLOps и production в DS-исследованиях 2.0/3.0 (2023/2024); Яндекс / МФТИ — «Машинное обучение и анализ данных» (2020)

Публикация: CADFEM Review (2020) — калибровка DEM-модели технологической массы зерновых. Полный список: portfolio.dmitryivanov.dev/ru/publications

Английский: B1–B2 — техническая документация и переписка; разговорный — в процессе. **Русский:** родной.