

DANIIL GONCHAR

ML Engineer / Data Scientist

5+ YEARS IN PRODUCTION ML

+7 (911) 600-27-65 / gonchardaniil1998@gmail.com / @gonchardaniil / github.com/Lainterus1

01 / PROFILE

ML Engineer / Data Scientist with 5+ years of experience. Over the past three years, I have focused on RAG, document processing, and AI agents. I handle system design, experimentation, quality evaluation, and MLOps handoff.

02 / EXPERIENCE

BANK DOM.RF

September 2023 - Present

ML Engineer

- **CorpGPT.** Designed a RAG system for Confluence, DWH, SQL data marts, PDFs, and scans with BM25 + bge-m3, Qdrant, reranking, version and permission filters, and source checks. Hallucination rate fell from ~18% to ~11%.
- **Document RAG.** Added table/page-aware chunking plus domain, version, and effective-date filters. Recall@10 rose from 0.72 to 0.88 on a corpus of 6,000-8,000 documents.
- **SQL Review Agent.** Built a LangGraph workflow to analyze SQL diffs and dependency graphs. Validation ran through a simulation layer with no direct DWH execution, cutting analysis time from 90 to 35 minutes.
- **RCA Agent.** Built a stateful agent with tools for logs, SLAs, lineage, and release data. Capped steps and retries and added early stopping; MTTR fell from 4.5 to 2.1 hours.
- **MLOps handoff.** Prepared API/Docker packages, schemas, evaluation sets, and documentation for CorpGPT and the SQL agent.

PIK GROUP

June 2022 - September 2023

ML Engineer

- **Service Desk.** Trained a multi-label RuBERT classifier with per-class thresholds, critical-incident rules, and fallback logic. Macro-F1 reached ~0.88; routing errors fell by 35-40%.
- **Construction documents.** Built a layout parsing + OCR + regex + RuRoBERTa/CRF pipeline with reference-data normalization. NER Macro-F1 reached ~0.86-0.87; about 70% of documents were processed automatically.
- **FESCO logistics.** Designed a Transformer for chain reconstruction and ETA prediction, with graph-based route constraints and ONNX/INT8 inference. At 2.5M+ events/day, manual corrections fell from ~70% to ~15%, while latency stayed below 500 ms.

1C-RARUS

February 2021 - June 2022

ML Engineer

- **B2B scoring.** Built a LightGBM model on CRM, transaction, RFM, and text features, using time-based validation, PSI monitoring, and scheduled retraining. An A/B test showed a 10.5% conversion lift; ROC-AUC was ~0.78-0.81.
- **Service Desk.** Built a CPU inference pipeline with TF-IDF + LightGBM, regex, dictionaries, and fuzzy matching. Macro-F1 reached ~0.84; ~70-75% of requests were routed automatically with latency below 50 ms.

03 / SKILLS

- **LLM / RAG:** Qwen3-Next, Qwen2.5, LangGraph, hybrid retrieval, bge-m3, bge-reranker-large, Qdrant.
- **AI Agents / Quality:** ReAct, tool calling, LLM-as-a-judge, Phoenix, Pydantic/JSON, guardrails, source attribution.
- **Document Intelligence:** Docling, EasyOCR, table-aware chunking, OCR, NER, RuRoBERTa/CRF.
- **Production / Data:** FastAPI, Kafka, PySpark, MLflow, Docker, Kubernetes, ONNX Runtime, PostgreSQL.
- **Core:** Python, SQL, PyTorch, Transformers, Classic ML.

04 / EDUCATION

The Bonch-Bruевич St Petersburg State University of Telecommunications, 2021

Higher education in Electronics and Nanoelectronics, Faculty of Fundamental Training.

05 / LANGUAGES

Russian - native | English - C1