Олег Жолобов

AI Engineer

olegzholobov.com · tffriend015@gmail.com · Telegram @sdflux

О СЕБЕ

AI-инженер с опытом разработки ML-систем более 4 лет. Специализируюсь на LLM-агентах, дообучении языковых моделей и корпоративном NLP.

ОПЫТ РАБОТЫ

viamobi

Июл 2025 - настоящее время

AI Engineer

- Спроектировал и внедрил в продакшен диалогового LLM-агента для монетизации контента; конверсия в покупку в чатах с предложениями продаж **на ~30% выше**, чем у операторов.
- Медианное время ответа **на 84% ниже**, а доля возвратных чатов **на 3,2 п.п. выше**, чем у операторов.
- Разработал многошаговые сценарии продаж с распознаванием интентов, учётом контекста и адаптивными стратегиями диалога.
- Дообучил локальную **Qwen** с **SFT и GRPO** для правильного порядка вызовов инструментов и сокращения лишних вызовов. Применил Schema-Guided Reasoning (SGR), достигнув **95% точности tool calling**.
- Разработал модульную обёртку над **LangGraph** для ускорения разработки агентов и управления версиями графов, узлами, маршрутизацией и промптами.

CIARS

Апр 2022 - июл 2025

ML-инженер · Руководитель ML-команды · Руководитель AI-направления

- Спроектировал **GraphRAG** с векторным и полнотекстовым поиском и рассуждениями по графу; достиг **98% recall и 96% precision** на тестах корпоративной базы знаний.
- Дообучил **Llama / Mistral** с **PEFT и LoRA**, применив распределённое обучение на **кластере из 8 GPU**.
- Разработал мультиагентные LLM-пайплайны для классификации и суммаризации документов и ответов на вопросы.
- Руководил командами **до трёх ML- и data-инженеров**: отвечал за технические решения, наставничество и внедрение моделей.
- Дообучил **BERT** для классификации интентов по каталогу товаров маркетплейса.

НАВЫКИ

LLM и агенты: RAG · LangGraph · LangChain · OpenAI API · Context engineering · Prompt engineering · MCP · LLM evaluation · LLM-as-a-judge · Langfuse · ElevenLabs · LlamaIndex

Машинное обучение: PyTorch · CatBoost · statsmodels · scikit-learn · TRL · Accelerate · JAX · PEFT · LoRA · Unsloth

Backend: Python · SQL · PostgreSQL · pgvector · Qdrant · FAISS · Elasticsearch · ClickHouse · Redis · Airflow

MLOps: Docker · Kubernetes · AWS · GCP · vLLM · GitLab CI/CD · MLflow · Prometheus · Grafana

Разработка с ИИ: Claude Code / Codex · Разработка по спецификациям и ревью кода

ОБРАЗОВАНИЕ И ЯЗЫКИ

Финансовый университет при Правительстве Российской Федерации 2026

Прикладная математика и информатика

Английский - C1 · Русский - родной