

Карпенко Сергей

Data Scientist (NLP/LLM)

Москва | +7 (962) 449-4951 | iskarp35@gmail.com
GitHub: github.com/SergeyKarpenko1 | Telegram: @Karpenko_Sergey1

Профессиональный профиль

Data Scientist с 3-летним опытом в области машинного обучения и NLP. Специализируюсь на разработке и внедрении решений на основе трансформеров и LLM. Имею успешный опыт реализации RAG-систем, оптимизации инференса моделей и интеграции ML-решений в бизнес-процессы. Достиг значительных улучшений в эффективности систем (до 20% повышение точности) и бизнес-показателях (рост прибыли на 14%).

Опыт работы

Data Scientist (NLP/RAG) | FronTech AI, Москва

Сентябрь 2024 — Февраль 2025

- Спроектировал и разработал систему Retrieval-Augmented Generation (RAG) для корпоративных данных
- Реализовал полный пайплайн обработки документов с 95% точностью извлечения текста
- Оптимизировал поиск релевантной информации, повысив точность на 20% через MMR
- Интегрировал GPT-4-mini для генерации ответов и решения NLP задач
- Построил масштабируемые цепочки обработки данных на LangChain

Стек: Python, PyTorch, Hugging Face, Transformers, OpenAI, FastAPI, LangChain, Chroma

Data Scientist (ML) | TWINO, Москва

Ноябрь 2023 — Август 2024

- Реализовал полный цикл разработки ML-моделей для финансового сектора
- Оптимизировал существующие модели для повышения точности предсказаний
- Разработал инструменты автоматизации ML-процессов
- Провел рефакторинг и оптимизацию кодовой базы

Стек: Python, Pandas, Sklearn, CatBoost, PostgreSQL, Git

Data Scientist/ML Engineer | XFIT, Ставрополь

Сентябрь 2021 — Май 2023

- Разработал и внедрил телеграм-бот на основе Sentence Transformers
- Создал систему прогнозирования бизнес-метрик с использованием ML
- Реализовал CV-решения для анализа посещаемости (YOLO, detectron2)
- Увеличил прибыль департамента на 14% через оптимизацию процессов

Стек: Python, PyTorch, Sklearn, XGBoost, CatBoost, Computer Vision

Технические навыки

- ML/DL:** PyTorch, Transformers (BERT, GPT), RAG, PEFT, Neural Networks
- Классический ML:** sklearn, CatBoost, XGBoost, LightGBM
- Data Processing:** Pandas, NumPy, PostgreSQL
- DevOps:** Git, Docker
- Frameworks:** FastAPI, LangChain
- Computer Vision:** YOLO, detectron2

Образование

Ставропольский государственный университет (2008)

Информатика и вычислительная техника, Автоматизированные системы обработки информации и управления

Языки

- Русский (родной)
- Английский (B1)

Дополнительная информация

- Активно слежу за развитием LLM и мультиагентных систем
- Имею опыт оптимизации инференса языковых моделей
- Участвую в ML-сообществах и конференциях
- Увлекаюсь сноубордингом и спортом