

КАРПЕНКО СЕРГЕЙ

DATA SCIENTIST



+7 962 449 49 51



github.com/SergeyKarpenko1



iskarp35@gmail.com



Karpenko_Sergey1



О СЕБЕ

Я Data Scientist **с 3-летним опытом** работы, специализируюсь на задачах обработки естественного языка (NLP) с применением нейросетей и трансформеров.

Реализовывал системы Retrieval-Augmented Generation (RAG) для интеграции внешних данных в генерацию текстов. Работал с трансформерами (BERT, GPT, PEFT), создавал и оптимизировал модели для классификации, предсказания и генерации текстов. Имею опыт использования AI-агентов (CrewAI) для построения автономных решений.

Участвовал в полном цикле разработки — от формулировки задач до внедрения и оценки результатов. Эффективно работаю в команде и самостоятельно. Вне работы увлекаюсь спортом и сноубордингом.



НАВЫКИ

- PYTORCH, BERT, GPT, RAG, PEFT
- HUGGING FACE, TRANSFORMERS
- FASTAPI, LANGCHAIN, CHROMA
- PYTHON, PANDAS, NUMPY, MATPLOTLIB, SEABORN
- SKLEARN, XGBOOST, CATBOOST, LIGHTGBM
- AIOGRAM.

TOOLS:

- GIT, SLACK, BASH
- SPARK, POSTGRESSQL
- DOCKER



ОБРАЗОВАНИЕ

СКФУ, ВЫСШЕЕ 2008

Автоматизированные системы обработки информации и управления

ELBRUS BOOTCAMP 2023

ПОВЫШЕНИЕ КВАЛИФИКАЦИИ DATA SCIENCE



ЯЗЫКИ

Английский - B1

Русский - родной

ОПЫТ РАБОТЫ



DATA SCIENTIST (NLP/RAG)

FronTech AI 2024 - 2025

Спроектировал и разработал с нуля систему **Retrieval-Augmented Generation (RAG)** для оптимизации поиска и генерации релевантной информации на основе корпоративных данных, включая регламенты, закупочную документацию и прайс-листы.

- Использовал библиотеки **PyPDFLoader** и **pytesseract** для обработки документов в различных форматах (PDF, изображения), обеспечив 95% точность извлечения текста.
- Реализовал семантическое разделение документов на чанки, что улучшило структуризацию данных и ускорило обработку.
- Настроил и оптимизировал базу данных **Chroma** для индексирования данных, что сократило время поиска релевантной информации.
- Применил алгоритмы **MMR (Maximal Marginal Relevance)**, что повысило точность поиска релевантной информации в векторной базе на 20%.
- Использовал **GPT-4o-mini** для генерации текстов и решения задач обработки естественного языка, что позволило сократить время и улучшить качество генерации ответов.
- Построил с использованием **LangChain** сложные цепочки обработки данных, автоматизировав ключевые сценарии работы с текстом.

Результат: Система успешно внедрена и обеспечивает быстрый и точный доступ к данным, что повышает эффективность сотрудников и снижает затраты на поиск и обработку информации.



DATA SCIENTIST

TWINO (Fintech) 2023 - 2024

- Участвовал в полном цикле разработки ML-моделей: от формулирования целей и задач до оценки экономического эффекта, включая сбор и подготовку данных, обучение, валидацию и тестирование моделей.
- Разрабатывал предложения и инструменты для улучшения процессов разработки ML-моделей, что позволило ускорить разработку и внедрение решений.
- Мониторил текущие ML-модели и оптимизировал их для повышения точности и эффективности.
- Инициировал и тестировал новые источники данных, что привело к улучшению ключевых метрик моделей.
- Рефакторил существующие R-скрипты и эндпоинты на Python для унификации и повышения производительности системы.



DATA SCIENTIST

X-Fit 2021 - 2023

- Выявлял проблемы в бизнесе, которые можно решить при помощи Data Science;
- обрабатывал и анализировал данные с использованием статистических методов (**python, pandas, numpy**).
- Создавал информативные графические представления данных и извлекал ключевую информацию, которая помогала в принятии решений (**matplotlib, seaborn**).
- Строил ML модели для оптимизации работы фитнес клуба, включая прогнозирование факторов, таких как посещаемость, выручка и эффективность маркетинговых мероприятий (**LogisticRegression**).
- Выявлял потребности клиентов в новых видах услуг при помощи анализа посещения различных зон клуба.
- Внедрял и оптимизировал модели, а также мониторил и улучшал их эффективность в рабочем процессе.

Достижения:

- Используя базу данных переписок отдела продаж с клиентами, разработал телеграм-бот с применением глубокого обучения (DL) и **Sentence Transformers (PyTorch)**, что помогло улучшить эффективность коммуникации отдела продаж с клиентами клуба.