



Ашихмина Надежда Анатольевна

email: ashikhmina.na@phystech.edu
телефон: +7 951 084 7362
tg: @Shikh_mi
GitHub: github.com/shikhmi

Желаемая должность

Стажёр Data Scientist (рекомендательные системы, Big Data)

Цель или краткое резюме (Summary)

Студентка 4 курса МФТИ. Имею практический опыт работы с данными: проектирование структур хранения, разработка SQL-запросов (DDL, DML), построение пайплайнов обработки и подготовки данных, автоматизация обработки в Python (pandas, numpy). Работала с PostgreSQL, настраивала процессы агрегации, фильтрации и анализа данных. Умею выстраивать полный цикл работы с данными: от проектирования структуры и загрузки до аналитики и валидации. Интересуюсь разработкой инфраструктурных решений и инструментов для работы с данными.

Образование

Московский физико-технический институт (МФТИ)

Прикладная математика и физика

2022 – н.в.

Beijing Institute of Technology (BIT)

Китайский язык и мультикультурная практика, Пекин

июль 2024

Опыт работы / практика

ML-инженер — ИНЭОС РАН, Москва

2024

- Провела полный цикл подготовки данных: очистка, нормализация данных, построение признаков и формирование обучающей выборки (134 000 образцов, 22 признака) (pandas, numpy).
- Разработала baseline-модель на XGBoost для предсказания структуры вещества по данным ЯМР-спектра, настроила пайпайн обучения и валидации (train/valid split).
- Достигла качества (метрика: Tanimoto similarity) = 0.65.

ML-researcher — Сколковский институт науки и технологий, Москва

2025 – 2026

- Подготовила обучающий датасет: агрегация, фильтрация, анализ выбросов, работа с большими массивами численных данных.
- Разрабатывала и обучала ML-модели, проводила сравнение подходов по метрикам MAE (энергия и силы).
- Формулировала и проверяла гипотезы по улучшению качества моделей.
- Автоматизировала пайпайн обработки данных и подготовки отчётов.

Проекты

Transformer-based Anecdote Generator (decoder-only LM)

2026

- Реализовала decoder-only Transformer в PyTorch (Byte-level BPE, Multi-Head Causal Self-Attention, GQA, ALiBi, RMSNorm, SwiGLU).
- Настроила полный pipeline обучения для задачи Causal Language Modeling (AdamW, Warmup + Linear Decay, teacher forcing).
- Обучила модель в течение 10 000 шагов.
- Достигла: Training Loss = 3.853, Validation Loss = 3.495 (Cross-Entropy).
- Обеспечила стабильную сходимость без выраженного переобучения.
- Реализовала autoregressive генерацию с temperature scaling и top-k sampling.

Система управления продажами (проект БД)

2024

- Спроектировала реляционную модель базы данных (ER-диаграмма, нормализация).
- Реализовала структуру БД в PostgreSQL (DDL: CREATE/ALTER/DROP).
- Разработала сложные SQL-запросы (DML: SELECT с JOIN, агрегаты, подзапросы).
- Реализовала представления, триггеры и хранимые процедуры.

Навыки

Технические

- Python (pandas, numpy, scikit-learn)
- PyTorch (нейросетевые модели, Transformer)
- SQL (JOIN, агрегаты, подзапросы), PostgreSQL
- Машинное обучение (классификация, регрессия, валидация, метрики)
- Глубокое обучение (нейросети, трансформеры)
- XGBoost
- Построение ML-пайплайнов
- Git, Linux

Дополнительно

- промпting и reasoning
- обработка и векторизация текста
- алгоритмы и структуры данных

Курсы

- ШАД МТС.
- Курс от Girafe AI. (МФТИ)
- Современный NLP. Большие языковые модели. (VK)
- Теория вероятности и статистика. (МФТИ)
- Алгоритмы и структуры данных. (МФТИ)

Соф-скиллы

- аналитическое мышление
- системность и аккуратность
- работа в команде
- стрессоустойчивость

Язык

- Русский — родной
- Английский — C1
- Китайский — B1

Достижения

- Призёр ВсОШ по химии (2021/2022)
- Победитель Сеченовской олимпиады (2021/2022)

Хобби

Фортепиано, боевые искусства, чтение русских классиков, кулинария.