

Дмитрий Балобин

Senior NLP / LLM Engineer

Москва, Россия

hanaan@internet.ru

t.me/d0rj3228

huggingface.co/d0rj

Опыт работы

Яндекс

Январь 2026 – Сейчас

Senior ML Engineer, Алиса на Поиске

- Оптимизация инфраструктуры обучения LLM на 32–64 GPU — ускорение обучения более чем в 2 раза (huggingface: yandex/AliceAI-T5-35B-A0.6B);
- Оптимизация модели в production — увеличение пропускной способности на 5,6% (RPS);
- Сокращение времени выполнения пайплайна оптимизации модели перед развёртыванием в production примерно на 40%;
- Исследование encoder-decoder и decoder-only MoE LLM архитектур.

Центральный Университет

Июнь 2026 – Сейчас

Ментор

- Поддержка развития студентов в обучении и карьере

Райффайзенбанк

Сентябрь 2025 – Январь 2026

Senior AI Engineer, CIB

- Разработка двух проектов с нуля до пилота за 4 месяца — AI Copilot и автономные агенты для онлайн-трейдинга;

2ГИС

Январь 2024 – Сентябрь 2025

Senior NLP Data Scientist

- Самостоятельная разработка и внедрение суммаризации отзывов (фича «Люди говорят») — повышение конверсии в клики;
- Разработка генеративной суммаризации для онлайн-поиска;
- Разработка матчинга товаров на теги — macro F1 = 0.97 на выборке с ручной разметкой;
- Проведение технических собеседований в свою и смежную команды;

Сбер

Июнь 2022 – Декабрь 2023

ML Engineer, Команда NLP B2C

- Разработка RLHF шага обучения для LLM GigaChat, проведение исследовательских экспериментов;
- Создание пилотов внедрения GigaChat в B2C и внутренние продукты банка. Создание retrieval части для бизнес кейсов RAG с помощью GigaChat;
- Доработка пайплайна обучения SFT GigaChat, адаптация внутренних моделей;

Сбер

Декабрь 2021 – Май 2022

Стажёр Data Scientist, Команда NLP B2C

- Речевая аналитика телефонных звонков (кластеризация, визуализация);
- Улучшение пайплайна обучения модели intent recognition, адаптация его под внутренние требования качества;

Образование

МФТИ ФПМИ

2025 – 2028

Аспирантура: Компьютерные науки и информатика

МФТИ ФПМИ

2023 – 2025

Магистратура: ИИТ, Машинное обучение и анализ данных

РТУ МИРЭА

2019 – 2023

Бакалавриат: Прикладная математика, Анализ данных

Статьи

Балобин Д. Ю., Сомов О. Д. Улучшение методов детекции сгенерированных текстов. – 2025.

Используемые технологии

Инструменты: PyTorch, Transformers, DeepSpeed, verl, vllm, langchain, PydanticAI, langfuse, k8s, FastAPI.