

Балобин Дмитрий

NLP ML Engineer

Москва, Россия

hanaan@internet.ru
t.me/d0rj3228
huggingface.co/d0rj

Опыт работы

Сбер	Июнь 2022 – По настоящее время
Junior Data Scientist	13 мес.
Команда NLP B2C	
<ul style="list-style-type: none">Разработка и отладка RLHF шага обучения GigaChat, проведение экспериментов;Доработка пайплайна обучения SFT GigaChat, адаптация внутренних моделей;Добавление в пайплайн обучения LLM логирования в ClearML и расчёта метрик качества;Разработка модели Intent Recognition для диалоговой системы (классификация текста, similarity search, rule-based NER). Увеличил количество интенгов одновременно с улучшением метрик на 3.5%;Сбор и валидация датасетов для модели классификации (Toloka, ПРОМ данные, opensource данные). Объём данных увеличился в 4 раза с 10000 до 40000 примеров;Оптимизация моделей для использования в ПРОМе (машины с CPU ресурсами). Ускорил общий инференс на целевом железе в 2 раза;Внедрение системы версионирования данных и отслеживания экспериментов ClearML в команде	
Сбер	Декабрь 2021 – Май 2022
Стажёр Data Scientist	6 мес.
Команда NLP B2C	
<ul style="list-style-type: none">Речевая аналитика телефонных звонков (кластеризация, визуализация);Улучшение пайплайна обучения модели intent recognition (перенос на PyTorch Lightning), адаптация его под внутренние требования качества;Увеличение скорости и качества на ПРОМе путём анализа существующих решений;Разработка Telegram ботов для оптимизации рабочих задач (проверка диалоговой логики, тестирование моделей)	

Образование

РТУ МИРЭА	2019 – 2023
Прикладная математика, Анализ данных	
СберУниверситет	Ноя 2022 — Янв 2023
Курс - Архитектура нейронных сетей: Трансформеры	
СберУниверситет	Янв 2022 — Май 2022
Курс - NLP (углубленный курс)	
Tinkoff Fintech	Фев 2021 — Май 2021
Курс - Разработка на Scala	

Используемые технологии

Языки программирования: Python;
Фреймворки: PyTorch, DeepSpeed, Megatron, Transformers, Sklearn, Streamlit, faiss, ONNX;
Инструменты: Git, DVC, Docker, MLFlow, Tensorboard, ClearML, ONNX, Poetry.