

# СЕНАТОВ ПЕТР

+7-(965)-586-79-36 | [senatoroff.p@gmail.com](mailto:senatoroff.p@gmail.com) | [github.com/PETEPEtrek](https://github.com/PETEPEtrek)

## ОБРАЗОВАНИЕ

### МФТИ

ФПМИ(ФИБТ) Информатика и вычислительные технологии

Долгопрудный, Московская область

Сентябрь 2020 – Июнь 2024

## ОПЫТ

- Data Science поток — работал с алгоритмами классического мл (регрессии, бустинги(градиентный, CatBoost, Xgboost), решающие деревья, итд), нейросетями (в частности для сегментации, детектирования изображений, то есть encoder-decoder), работал с моделями в области NLP (BERT, GPT) и рекуррентными нейросетями(RNN, LSTM, GRU)), а также метриками итд.
- На курсе статистики изучал как работать с данными, в каких случаях стоит строить гипотезы и какими критериями их проверять, также прошел курс по А/Б тестированию (строил свои А/В тесты с применением стратификации, CUPED, линеаризации итд.) и Casual Inference.
- На Стартап-треке участвовал в проекте wikiuniversity, где работал в разработке пайплайна для преобразования документов в md формат с использованием LayoutParser для разбития документа на сущности (картинки, текст, заголовки ...), Tesseract для распознавания текст (дообучался на своих наборах), и docker для деплоя
- В рамках дипломной работы экспериментировал (искал в научных статьях + придумывал свои) с оффлайн метриками для дообученных на persona-chat больших языковых моделей (7B) для исследования корреляции с 'персонофицированностью' модели
- Также на курсе распределенных систем проходил MPI, Cuda, Hadoop, Spark, Hive. На курсе по базам данных работал с PostgreSQL
- В рамках университетских курсов:C++ - реализовал свой вариант таких STL-контейнеров, как std::vector, std::list, std::unordered set, а также std::unique/shared ptr, std::optional etc., Операционные системы(системные вызовы, сокеты, каналы, epoll, etc), Введение в анализ данных, курс теории и практики многопоточной синхронизации, небольшие проекты на golang
- Уровень знания английского Intermediate

## YANDEX

Команда контекста (NLP) | *Python, YT, YQL, Nirvana*

Июль 2023 – Октябрь 2023

Работал над улучшением качества браузерного перевода:

- Написал тулзу для сбора параллельного html корпуса (предложения и параграфы)
- Собрал параллельные корпуса из предложений и параграфов с html разметкой и провел серию экспериментов с использованием браузерных моделей перевода, используя свои корпуса
- Получил прирост в  $\sim 0.7 - 1.5 BLEU$  на моделях со своими данными, а также статзначимые победы моих моделей на SbS и Pointwise тестах, модели отправились в релиз

## TINKOFF

Команда контроля качества | *Python, SQL*

Декабрь 2023 – настоящее время

- Построил процесс сбора CSAT представителей, а также последующего их перевода на тестовый период и блок
- Участвовал в разработке и релизе новых опросников
- Писал автоматические нарушения представителей
- Администрировал создание нового личного кабинета для внутреннего пользования

## TECHNICAL SKILLS

**Languages:** Python, C/C++, Postgres (SQL), Go

**Developer Tools:** Git, Docker, VS Code, Visual Studio, PyCharm, IntelliJ, Clion, GoLand, DataGrip, DataSpell

**Libraries:** pandas, NumPy, Matplotlib, sklearn, nltk, opencv, seaborn, scipy, PyTorch, Tensorflow