

АЛЕКСАНДРА ШИКУНОВА

8 929 980 69 03 | [email](#) | [website](#) | [github](#) | [hse.ru](#)

ОБРАЗОВАНИЕ

НИУ ВШЭ

Магистратура “Лингвистическая теория и описание языка”

Москва

2024 – сейчас

НИУ ВШЭ

Бакалавриат “Фундаментальная и компьютерная лингвистика”, майнор “Интеллектуальный анализ данных” 2020 – 2024

Москва

ОПЫТ РАБОТЫ

NLP-инженер

ПАО Сбербанк, центр ИИ блока “Сервисы”

октябрь 2023 – сейчас

Москва

- для ускорения обработки новостного потока для департамента кибербезопасности дедублицировала, кластеризовала и суммаризировала новости
- обучила SFT SLM с 400M параметров суммаризировать новости в 17 раз быстрее 7B модели при сравнимом качестве, удешевив внедрение в прод
- построила единый пайплайн для задач инфопоиска центра снабжения, тестирующий несколько подходов (tf-idf, BM25, эмбеддеры)
- свела задачи NER и поиска по реестру товаров для центра снабжения к regex, на продуктивизации кода для поиска сопровождала стажёра

Аналитик больших данных

ПАО Мегафон, команда чат-бота

июнь 2022 – октябрь 2023

Москва

- регулярное обновление продовых моделей intent & entity recognition
- обернула модель кластеризации для разведки новых интенгов в приложение для ассессоров

Стажёр-исследователь

НУЛ по формальным моделям в лингвистике, НИУ ВШЭ

2021 – сейчас

Москва

- Опубликовала 5 статей по теоретической лингвистике, сделала 10+ докладов на международных конференциях
- Участвовала в лингвистических экспедициях в мокшанский, севернохантыйский и лесной ненецкий языки

Учебный ассистент

НИУ ВШЭ

2021 – 2023

Москва

- Проассистировала курсы дискретной математики (2021), Python (2022), теории вероятности и матстатистики (2023)
- Проверка ДЗ и проверочных работ, проведение консультаций со студентами бакалавриата

ЛИЧНЫЕ ПРОЕКТЫ

Обработка данных и инфографики для фонологии малых языков России | Python, FastAPI, Praat

2023 — сейчас

- Проверила лексико-статистическую гипотезу о продуктивности некоторых правил мокшанской фонологии
- Собрала, разметила и обработала аудиозаписи на лесном ненецком
- По данным можно строить инфографики через запрос к веб-сервису

Nonce RNN | pytorch, wandb

2023

- Провела эксперименты по генерации псевдослов по словарю для 20+ языков

MessageHDD | Python

2022

- Создала и протестировала метрику лексического разнообразия для корпуса коротких текстов

ТЕХНИЧЕСКИЕ НАВЫКИ

Языки программирования: Python, SQL, HTML/CSS, R

Фреймворки & библиотеки: Huggingface, Pytorch, sklearn, pandas, numpy, nltk, PySpark, FastAPI, Flask, streamlit, wandb, Praat

Иностранные языки: русский родной, английский C2 (TOEFL 2023), французский B1, немецкий A2, персидский A2