



Михаил Васильев

Deep learning engineer

Опыт

2023—н.в. **Старший специалист по машинному обучению, Макс Групп, Москва**

Проект: создание комплексного решения для обеспечения безопасности в корпоративной сети на основе неструктурированных данных.

Инструменты: python, transformers, EfficientNet, MobileNet, YOLO, OpenCV, PIL, PyOD, pandas, sklearn, pytorch, lightning, numpy, matplotlib, plotly, huggingface, onnx, fastapi, uvicorn, pyinstaller, pywin32, optimum, airflow, mlflow, cvat, natasha, deeppavlov, BERT, whisper, Ollama.

- реализовал нейросетевой модуль для поиска нарушений закона о персональных данных, количество детектируемых классов увеличено с 14 до 36, accuracy top 1 увеличена до 98.9
- подготовил модуль для анализа содержимого отсканированных документов: поиск текста, таблиц, печатей, подписей и корпоративных бланков, количество классов увеличено с 5 до 19, mAP@.5 улучшен с .89 до .94
- реализовал поиск чувствительных данных в текстовых файлах, добавил модуль NER
- создал ансамбль алгоритмов для поиска аномалий на табличных данных, в том числе на временных рядах
- реализовал поиск чувствительных данных в аудио-файлах
- собрал и организовал разметку 8 датасетов для задач классификации и object detection

* 25 апреля 1987

☎ +7 (916) 198 81 83 • ✉ gnu.xinm@gmail.com

🌐 onixlas.github.io • **in** [michael-vasiliev-ds](#)

🔄 [onixlas](#) • 📍 [LaHundo](#)

Пет-проекты

2024 **Тим-лид и технический эксперт, CheckDocAI, Москва**
Проект: Телеграм-бот с ИИ модулем для контроля качества оформления документов для ООО «Гольфстрим»

Инструменты: aiogram, yolo, onnx, albumentations, cvat

- руководил командой из двух дата-сайнтистов и бекенд-разработчика
- проект завершён и внедрён в коммерческую эксплуатацию
- ежемесячная экономия — 40 человеко-часов

Хакатоны

2024 **VK HSE Data Hack, 1 место, Москва**

Хакатон по классификации новостных статей на 21 класс. В нашем решении комбинируются результаты работы небольшого классификатора на базе трансформерной архитектуры и предсказания LLM

Инструменты: transformers, Saiga3 8b, taiga dataset, streamlit

- обогатил датасет
- подобрал zero-shot classification модель
- обучил модель-классификатор
- обеспечил координацию работы команды
- презентовал результаты

Выступления

25.06.2024 **Опыт обучения и применения нейросетей в качестве модуля российской DСАР системы, Moscow Python Meetup, Москва, Компания Makves (входит в группу компаний «Гарда»)** разрабатывает российскую DСАР (data-centric audit and protection) систему для защиты корпоративных данных. Для анализа неструктурированных данных необходимо применять нейросети. В докладе я рассказал о проблемах, с которыми мы столкнулись при создании нейросетей, от этапа сбора и разметки данных и до создания нескольких микросервисов.

* 25 апреля 1987

☎ +7 (916) 198 81 83 • ✉ gnu.xinm@gmail.com

🌐 onixlas.github.io • **in** [michael-vasiliev-ds](#)

🔄 [onixlas](#) • 📍 [LaHundo](#)

Образование

- 2024 **Анализ данных на языке SQL**, УЦ «Специалист», Москва, повышение квалификации
- 2022—2023 **Computer Vision Engineer**, *Deep Learning School ФПМИ МФТИ*, Москва, профессиональная переподготовка
- 2022 **Специалист по Data Science**, *Яндекс Практикум*, Москва, профессиональная переподготовка
- 2021—2022 **Введение в искусственный интеллект и нейросети для авиационных приложений**, *МАИ*, Москва, повышение квалификации
- 2005—2008 **Перевод и переводоведение**, *МАИ*, Москва, специалитет
- 2003—2009 **Авиационная и ракетно-космическая теплотехника**, *МАИ*, Москва, специалитет

Языки

Русский	■■■■■	родной
Английский	■■■■■	B2
Немецкий	■■■■■	B2
Эсперанто	■■■■■	B2

Навыки и технологии

- Deep Learning
- NLP, NER
- Computer Vision
- Speech Recognition
- Machine learning
- Anomaly Detection
- Data analysis
- Data visualisation
- Statistics
- Python
- SQL
- Linux
- Docker
- YOLO
- Natasha
- ONNX
- HuggingFace
- Ollama
- U-Net
- AirFlow
- MLFlow
- CVAT
- Plotly
- Pytorch
- Lightning
- Pandas
- NumPy
- Sklearn
- FastAPI
- uvicorn
- PyOD
- PySAD
- Optimum
- pywin32
- CatBoost
- XGBoost
- PostgreSQL

* 25 апреля 1987