



Михаил Васильев

Deep learning engineer

Опыт

- 2025–н.в. **Старший инженер по машинному обучению, Райффайзен Банк**
- 2023–2025 **Старший специалист по машинному обучению, Makerves**

Проект: разработка и внедрение RAG-системы

- Разработал и внедрил RAG-систему для автоматизации обработки запросов заказчиков
- Оптимизировал гиперпараметры системы с использованием библиотеки Ragas и языковой модели GigaChat

Инструменты: LangChain, Ollama, Saiga, GigaChat, HuggingFace, FastAPI, Ragas

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

- 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, подпiseй 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
- Собрал и организовал разметку 8 датасетов для задач классификации и object detection

Инструменты: transformers, YOLO, PyOD, Pandas, Sklearn, PyTorch, lightning, NumPy, HuggingFace, ONNX, FastAPI, uvicorn, PyInstaller, optimum, CatBoost, CVAT, natasha

Пет-проекты

2024 Тим-лид и технический эксперт, *CheckDocAI*

Проект: Телеграм-бот с ИИ модулем для контроля качества оформления документов для ООО «Гольфстрим», позволяет значительно сократить время на проверку и улучшить точность

- Руководил командой из двух data-сайентистов и бекендер разработчика, отвечал за разработку и внедрение проекта
- Проект успешно внедрён в коммерческую эксплуатацию, ежемесячная экономия — 40 человеко-часов

Инструменты: aiogram, YOLO, ONNX, Albumentations, CVAT

Хакатоны

2024 VK HSE Data Hack, 1 место

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

- Обогатил датасет
- Подобрал zero-shot classification модель
- Обучил модель-классификатор
- Обеспечил координацию работы команды
- Презентовал результаты

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

Доклады

24.05.2025 Поиск аномалий с использованием Python: от теории к практике, *Positive Hack Days*

2025 Серия докладов: Поиск аномалий в данных, алгоритмы, *Moscow Python Meetup*

2024–2025 NLP и CV нейросети в защите данных: опыт Makves DCAP, *Moscow Python Meetup*

Образование

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

Языки

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

Навыки и технологии

Deep Learning & LLMs

- RAG, Prompt Engineering
- Qwen, Llama, GigaChat
- LangChain, Ollama, Ragas
- Transformers, BERT

Computer Vision

- YOLOv8, U-Net
- OpenCV, CVAT
- Albumentations

Anomaly Detection

- PyOD, RRCF
- Isolation Forest, ECOD
- HBOS, PySAD

NLP & Speech

- NER (natasha)
- Whisper, HuBERT
- Text classification

Vector Search

- FAISS, Qdrant, Milvus
- sentence-transformers

MLOps & Deployment

- Docker, Linux, FastAPI
- MLflow, Airflow
- ONNX, PyInstaller

Testing & Reproducibility

- pytest
- Git, CI/CD basics
- Experiment tracking

Data & Core ML

- Python, SQL, pandas
- scikit-learn, CatBoost
- PostgreSQL, MySQL