



# Михаил Васильев

*Deep learning engineer*

## Опыт

2023—н.в. **Старший специалист по машинному обучению, Makves**

Проект: **разработка и внедрение RAG-системы**

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

- разработал и внедрил RAG-систему для автоматизации обработки запросов заказчиков
- оптимизировал гиперпараметры системы с использованием библиотеки Ragas и LLM GigaChat

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

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

- 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](mailto:gnu.xinm@gmail.com)

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

🌀 [onixlas](#) • 📍 [LaHundo](#)

## 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.

Инструменты: aiogram, YOLO, ONNX, Albumentations, CVAT

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

## 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

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

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

## Доклады

29.01.2025 **NLP нейросети в защите данных: опыт Makves DCAP, Moscow Python Meetup № 98**

25.06.2024 **Опыт обучения и применения нейросетей в качестве модуля российской DCAP системы, Moscow Python Meetup № 91**

\* 25 апреля 1987

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

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

🔄 onixlas • 📍 LaHundo

2/3

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

\* 25 апреля 1987