



# Mikhail Vasiliev

*Deep Learning Engineer*

## Experience

2025—  
present **Senior Machine Learning Engineer, Raiffeisenbank**

2023—2025 **Senior Machine Learning Specialist, Makves**

Project: *Development and implementation of RAG system*

- Developed and implemented a RAG system for automating customer request processing
- Optimized system hyperparameters using Ragas library and GigaChat LLM

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

Project: *Creation of a comprehensive security solution for corporate networks based on unstructured data*

- Implemented a neural network module for detecting violations of personal data laws, increasing detectable classes from 14 to 36 with top 1 accuracy reaching 98.9
- Developed a module for analyzing scanned document content: text, tables, stamps, signatures and corporate forms detection, increasing classes from 5 to 19 with mAP@.5 improved from .89 to .94
- Implemented sensitive data detection in text files with added NER module
- Created an ensemble of algorithms for anomaly detection in tabular data, including time series
- Developed sensitive data detection in audio files
- Collected and organized labeling for 8 datasets for classification and object detection tasks

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

## Personal Projects

### 2024 **Team Lead and Technical Expert**, *CheckDocAI*

Project: *Telegram bot with AI module for document quality control for Gulfstream LLC, significantly reducing verification time and improving accuracy*

- Led a team of two data scientists and a backend developer, responsible for project development and implementation
- Successfully deployed for commercial use with monthly savings of 40 man-hours

Tools: aiogram, YOLO, ONNX, Albumentations, CVAT

## Hackathons

### 2024 **VK HSE Data Hack**, *1st place*

Hackathon for news article classification into 21 categories. Our solution combined results from a small transformer-based classifier and LLM predictions

- Enriched the dataset
- Selected zero-shot classification model
- Trained classifier model
- Coordinated team work
- Presented results

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

## Talks

24.05.2025 **Anomaly Detection with Python: from theory to practice**, *Positive Hack Days*

2025 **Lecture Series: Anomaly Detection in Data, Algorithms**, *Moscow Python Meetup*

2024—2025 **NLP and CV Neural Networks in Data Protection: Makes DCAP Experience**, *Moscow Python Meetup*

## Education

- 2024 **Data Analysis with SQL**, Training Center “Specialist”, professional development
- 2022—2023 **Computer Vision Engineer**, Deep Learning School, MIPT, professional retraining
- 2022 **Data Science Specialist**, Yandex Practicum, professional retraining
- 2021—2022 **Introduction to AI and Neural Networks for Aviation Applications**, MAI, professional development
- 2005—2008 **Translation and Translation Studies**, MAI, specialist degree
- 2003—2009 **Aviation and Space Thermal Engineering**, MAI, specialist degree

## Languages

Russian	■■■■■	native
English	■■■■■	B2
German	■■■■■	B2
Esperanto	■■■■■	B2

## Skills and Technologies

- |                      |               |              |
|----------------------|---------------|--------------|
| ○ 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      | ○ MySQL      |