

# Olga Bystrova

SENIOR DATA SCIENTIST · NLP LEAD

Moscow, Russia

✉ [bystrovaolgavl@gmail.com](mailto:bystrovaolgavl@gmail.com) | [ovbystrova](https://www.linkedin.com/company/ovbystrova) | [ovbystrova](https://www.linkedin.com/company/ovbystrova)

## Work Experience

### Sber

Moscow, Russia

SENIOR NLP DATA SCIENTIST

Aug. 2021 - Now

- **Deputy Team Lead** for 8 data scientists: team management, stakeholder communication, performance evaluation, candidate interviews
- **Lead DS** in project for extracting information from contracts and additional documentation: 5 000 docs per month in production
- Used **data augmentation** and **active learning** approaches to speed up annotation process in several projects
- Mentored 3 team members

### Sber

Moscow, Russia

DATA SCIENTIST

Aug. 2020 - Aug. 2021

- Build and deployed **an ensemble of NER models** for contract recognition system
- Worked on comprehensive **reviews system** from internal clients: sentiment analysis, topic modeling, custom classification models
- Improved process of documents's report generation with **summarization model** as a first step

### Visiology

Moscow, Russia

JUNIOR DATA SCIENTIST

Sep. 2019 - Jun. 2020

- Virtual assistant ViTalk team: improved assistant's functionality to work with relative dates written by user
- Worked on speech2text algorithm that allows user to control BI dashboards by voice

### Higher School of Economics — National Research University

Moscow, Russia

TEACHER ASSISTANT

Jan. 2020 - Jan. 2021

- Assisted on "Introduction to Data Analysis", "Introduction to Deep Learning" courses, Faculty of Computer Science

## Skills

<b>Scientific Areas</b>	DL, NLP: NER, Summarization, QA, Text Generation, Audio: ASR
<b>Deep Learning and Data Analysis</b>	PyTorch, Sklearn, Numpy, Pandas, Transformers, NLTK, Torchtext, Spacy, natasha, SQL (basic), Vosk
<b>Data Science workflow</b>	Git, Catalyst, Pytorch-lightning, Wandb, Confluence, Jira, Docker (basic)
<b>Languages</b>	Russian (native), English (advanced), French, German

## Projects

### InstructionNER: Generative NER

[github](https://github.com/ovbystrova/instructionner)

[HTTPS://GITHUB.COM/OVBYSTROVA/INSTRUCTIONNER](https://github.com/ovbystrova/instructionner)

July, 2022

- Unofficial implementation of InstructionNER paper
- Tuned T5 model to extract entities from input text

### Neural Fake News Detector and text GANs

[HSE](https://github.com/ovbystrova/dpl)

[HTTPS://GITHUB.COM/OVBYSTROVA/DPL](https://github.com/ovbystrova/dpl)

June, 2020

- Used GPT-2 model to generate short texts with different sampling types
- Built several (BiLSTM with different embeddings, BERT models) binary classifiers to detect fake texts
- Compared accuracy of classifiers based on sampling type
- Trained language models as GANs with explored classifiers

## Education

### HSE (Higher School of Economics)

Moscow, Russia

M.S. IN COMPUTATIONAL LINGUISTICS

Sep. 2020 - June. 2022

- GPA: 8.14
- Master Thesis: "NLP Augmentation Techniques for Russian Language".

### HSE (Higher School of Economics)

Moscow, Russia

B.S. IN FUNDAMENTAL AND APPLIED LINGUISTICS

Sep. 2016 - June. 2020

- GPA: 8.04
- Bachelor Thesis: "Neural Fake News Detector as a New Evaluation Metric for Text GANs".

## Certifications

---

2022 **Deep Learning Specialization**, deeplearning.ai  
2020 **Deep Learning in NLP**, DeepPavlov, MPTI  
2020 **Data Science Math Skills**, Duke University  
2020 **SQL Interactive Trainer**, Far Eastern Federal University

*Coursera*  
*Moscow, Russia*  
*Coursera*  
*Stepic*

## Publications

---

### Unreasonable Effectiveness of Rule-Based Heuristics in Solving Russian SuperGLUE Tasks

*Dialogue conference, Moscow*

COWRITER

*June 2021*

- <https://arxiv.org/abs/2105.01192>
- Implemented baseline solutions for RSG tasks: majority, random choice and random balanced choice
- Created heuristics for RUSSE, PARus, RuCoS tasks