

# Tatiana Sycheva

(eligible to work without restrictions)

## Data Scientist

Dublin, Ireland • owl0toto@gmail.com • [LinkedIn](#) | [GitHub](#)

Data Scientist Intern at Wildberries, Russia's largest online retailer and marketplace. Skilled in Python, deep learning, AI modeling, and cloud platforms such as Apache Spark and Hadoop. Experienced in data analysis with a strong aptitude for solving interdisciplinary challenges and quickly adapting to new technologies. MSc in Data Science (fully English-taught) from HSE University, graduating with honors.

## TECHNICAL SKILLS

**Languages:** Python, SQL

**Technologies:** Pandas, Scikit-Learn, Matplotlib, Seaborn, Numpy, TensorFlow, PyTorch, Keras, Markdown, LaTeX

**Tools:** GitHub, Google Colab, Jupyter, PyCharm, PostgreSQL, Apache Spark, Hadoop

## EDUCATION

**National Research University – Higher School of Economics**

September 2022 – July 2024

*Master's degree in Computer Science (with honors)*

- Applied Statistics, Linear Algebra in Applications, Information Theory and Combinatorial Theory of Search, Machine Learning and Data Mining, Deep Learning, Graph Algorithms, Large Scale Machine Learning, Modern Methods of Data Analysis, LLM, NLP

## PROFESSIONAL EXPERIENCE

**Data Scientist (remote)**

Wildberries, Horizontal ML, Computer Vision Team

March 2025 – now

- Focusing on Image-to-Text and Text-to-Image models and ViT for advanced image understanding and feature extraction. This includes fine-tuning deep learning models for tasks such as image captioning, multimodal embeddings and cross-modal retrieval, enhancing the interaction between textual and visual data.

**Data Scientist intern (remote)**

Wildberries, User Reputation Team

October 2024 – February 2025

- Contributed to a machine learning project focused on detecting obscene language in reviews and identifying veiled words
- Utilized LLM transformer models and fine-tuned BERT (RuBERT for the Russian language) for classification and NER tasks. Trained FastText and BiLSTM models and employed regular expressions
- Achieved the highest predictive performance among 32 candidates during the model evaluation

**Scientific Researcher**

JSC Sustainable Development

October 2017 – June 2021

- Wrote Python/pandas scripts for data mining and statistical analysis in biotechnology and sustainable development projects
- Analyzed large datasets on atmospheric emissions and heavy metal pollutants by industry, identifying key trends using Scikit-Learn, Matplotlib, and Excel
- Conducted hypothesis testing and presented results at research seminars, co-authoring articles on project methods and outcomes

## PROJECTS

**Graph Augmented Anisotropic Diffusion** | PyTorch, Graph Neural Networks, GitHub, LaTeX

October 2023 – June 2024

## National Research University – Higher School of Economics

- Enhanced Graph Neural Networks (GNNs) with anisotropic filters derived from Laplacians to improve Graph Anisotropic Diffusion (GAD) layers.
- Evaluated model performance on the ZINC dataset for predicting molecular solubility.

### **MuseGAN: Music Generation with GANs** | *Generative Adversarial Networks, NumPy, Deep Learning* **March 2024 – April 2024**

#### National Research University – Higher School of Economics

- Developed a generative model using GANs to produce multi-bar musical compositions, integrating melody, harmony, and rhythm
- Utilized PyTorch and NumPy for model development and data manipulation

### **BoolQ Question-Answering Project** | *NLP, Pretrained Embeddings (Word2Vec, FastText, BERT), Hugging Face* **February 2024**

#### National Research University – Higher School of Economics

- Developed a yes/no question-answering system using the BoolQ dataset from the SuperGLUE benchmark. Implemented various techniques such as traditional machine learning models (SVM, Logistic Regression) with Word2Vec and FastText embeddings, as well as fine-tuned BERT embeddings
- Compared the performance of models using accuracy metrics and provided insights for future improvements

### **Lane Detection Computer Vision Project (LaneNet, HNet)** | *OpenCV* **January 2024**

#### National Research University – Higher School of Economics

- Developed a lane detection solution using semantic segmentation, pixel embedding, and clustering algorithms for accurate lane identification.

### **Film Success Prediction Model** | *Python, Pandas, Scikit-Learn, Matplotlib* **November 2023**

- Built a predictive model using a dataset of 80,000+ films, leveraging regression models to predict film success based on features like genre, budget, and ratings.

## PUBLICATIONS AND AWARDS

- [doi.org/10.1016](https://doi.org/10.1016) (Scopus)
- [doi.org/10.1016/1](https://doi.org/10.1016/1) (Scopus)
- [doi.org/10.1088](https://doi.org/10.1088) (Scopus)
- [doi.org/10.3390](https://doi.org/10.3390) (Scopus)
- Prize-winner (3rd place) at the China-Russia Innovation and Entrepreneurship Competition, organized by Liuzhou New Silk Road Technology Transfer Institute (November 2020)