Andrei Tregubov

Data Scientist · R&D Engineer

Munich, Germany | andrey.tregubov2001@gmail.com
linkedin.com/in/andrei-tregubov-rnd | github.com/oster-field | andrei-tregubov.works
Born 31.03.2001 | Authorised to work full-time in Germany

PROFILE

Research-heavy Data Scientist and R&D Engineer with **3+ years** of experience building end-to-end data pipelines and **production-ready machine learning models** on large-scale time series data. Specialised in **signal processing,** and probabilistic **risk modelling** applied to **real-world data**. MSc in Computational Mathematics, with a **strong mathematical foundation** and a proven ability to deliver robust, deployable software. **Author of peer-reviewed research** published in Springer journals.

PROFESSIONAL EXPERIENCE

Data Scientist & R&D Engineer

Institute of Applied Physics | Jul 2021 – Nov 2024 · Remote

Worked in a small research cross-functional team, driving technical and analytical work.

- **Independently designed and built** an end-to-end sensor data processing application, reducing dataset processing time from 10 hours to 15 minutes (**40x speedup**), replacing fully manual, dataset-specific workflows with a fully automated solution covering **ingestion, quality control, feature extraction, and anomaly detection**
- **Engineered** a domain-driven **data processing methodology** to clean, segment, and extract representative statistical features from over **2.1 billion** time-series data points, enabling reliable **rare event analysis** from long-term field measurements
- **Developed** a **machine learning model** for **predictive risk assessment**, achieving **93.4% accuracy**, subsequently commercialised, **improving safety of maritime operations**
- **Formulated** a novel analytical model (formulated and proved a previously unknown mathematical theorem) for **probability estimation of rare events** in complex stochastic time-series data
- **Awarded a personal research grant**; presented findings at **19 international conferences**; results published in **2 Springer journals** (Scopus-indexed)

EDUCATION

MSc in Computational Mathematics · University of Passau · Passau, Germany · Sep 2023 – Mar 2026 · GPA 1.8
BSc in Fundamental Mathematics · Higher School of Economics · Moscow, Russia · Sep 2019 – Jun 2023 · GPA 1.6

SKILLS

Programming languages: Python, MATLAB, R, SQL

Python libraries: Pandas, NumPy, SciPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, PyTorch, XGBoost, Statsmodels, SQLAlchemy, Transformers

Tools: Git, Docker, AWS, LaTeX, Jupyter, AI-assisted development (ChatGPT, Claude, Copilot)

Core expertise: Machine Learning, Deep Learning, Research, Mathematics, Predictive Modelling, Data Engineering, Statistical Modelling, Anomaly Detection, Time-Series Analysis, Signal Processing, Big Data, Hypothesis Testing, Correlation Analysis, Probability Theory, Numerical Methods, Scientific Computing, Data Visualisation

RESEARCH & PUBLICATIONS

Frequency Spectra and Surface Wave Height Distributions Based on Measurements off Sakhalin Island
Atmospheric and Oceanic Physics · Springer (Scopus-indexed) · Oct 2025 · DOI: 10.1134/S0001433825700884
Large-scale analysis of real-world sensor data; statistical modelling of extreme event probability

Probability Distributions for Finite Ensembles of Irregular Waves
Springer Proceedings in Earth and Environmental Sciences (Scopus-indexed) · Nov 2023 · DOI: 10.1007/978-3-031-47851-2_19
Numerical simulation and probabilistic modelling of anomalous events in time-series data

LANGUAGES

German B2-C1 · English C2 · Russian C2 (Native)

Munich, March 2026