

30-383 Kraków, Poland

+48 792 608 016

yuvolkotrub@gmail.com

Holder of a long-term EU residence permit (Poland)

in yuriy-volkotrub 0000-0002-3114-3798 yuvolkotrub yvolkotr

# Yuriy Volkotrub

PhD Physicist and Data Scientist with 7+ years of professional experience in data science, machine learning, and large-scale experimental data analysis. Member of the [ATLAS](#) Collaboration at CERN, contributing to one of the world's largest international scientific collaborations. Applies advanced statistical methods and ML algorithms to high-dimensional datasets for [high-energy physics research](#).

## Skills

### Technical Skills

Python (NumPy, Pandas, Scikit-Learn, SciPy, Matplotlib, Seaborn), C++, SQL (PostgreSQL), *Mathematica*  
Docker, Google Cloud Platform, Azure (basics), Databricks (basics)  
PyTorch, Keras, Git, Jira, Confluence, L<sup>A</sup>T<sub>E</sub>X, Vim, Bash, ssh, OpeMPI (basics)

### Key Skills

- Statistical analysis and machine learning
- Technical presentations
- Ability to learn quickly and adapt to new technologies and methodologies
- Validation of prediction models
- Agile & Scrum
- Problem-solving

## Experience

- Nov. 2023 – **Applied Data Scientist, ATLAS Experiment @ CERN**, Jagiellonian University, Kraków, Poland.
- Jun. 2026
- Built and evaluated machine learning models (classification, regression and anomaly detection) on large-scale experimental datasets using Python (NumPy, SciPy, pandas, scikit-learn, PyTorch) to identify patterns in high-dimensional data for rare-decay and Higgs boson analyses.
  - Designed end-to-end analytical pipelines: ingesting heterogeneous data sources, cleaning and validating data, handling missing and implausible values, performing feature engineering, model selection and cross-validation, supported by extensive exploratory data analysis and visualisation.
  - Identified, collected and integrated data from multiple internal sources (distributed storage systems, metadata catalogues, Monte Carlo simulations), ensuring data quality, consistency and reproducibility of results.
  - Translated complex domain and research questions into analytical and machine-learning tasks, and communicated model results and limitations to non-ML experts within an international, data-driven team of ~50 physicists.
  - Developed and optimised multivariate classifiers for background rejection and signal extraction, and contributed to validation and calibration strategies analogous to model validation and monitoring in industry; [results set the most stringent ATLAS limits on Higgs self-coupling to date](#).
  - Co-authored a machine-learning-driven [study](#) on PET image reconstruction, applying probabilistic modelling, maximum-likelihood and Bayesian inference with PyTorch and MCMC to sparse, high-dimensional medical imaging data.
  - Followed good engineering practices: version control (Git), reproducible analysis pipelines and code reviews within the collaboration.
- Dec. 2023 – **Independent Data Science Projects, Freelance Work.**
- Present
- Applied statistical learning methods to business-oriented datasets, developing end-to-end data analysis and machine learning pipelines, including data preprocessing, feature engineering, model training, validation, and performance evaluation.

- Oct. 2021 – **Applied Data Scientist, ATLAS Experiment @ CERN**, AGH University of Kraków, Poland.
- Oct. 2023
- Trained and applied Boosted Decision Trees for signal–background classification and regression tasks on hardware-encoded parameters, improving particle identification (by 30%) and optimizing statistical significance in physics analysis.
  - Developed C++-based statistical models for background measurement in top-quark pair production.
  - Processed and analyzed terabytes of Monte Carlo simulation data, automating validation processes.
  - Experimental data analysis; analysis support in validation of data processing from experimental measurements.
  - Collaborated with an 8-person research team on data analysis for the ATLAS collaboration.
  - Performed advanced statistical analysis on a combination of the ATLAS and CMS results under the [STRONG-2020 project](#).
- Jun. 2021 – **Scientific Researcher/BAND Summer Fellowship, BAND Collaboration**, USA (remote).
- Sept. 2021
- The project involved remote collaboration with four institutions across the USA to pursue an accurate description of the properties of atomic nuclei and collisions between nuclei:
- Collaborated remotely with US institutions to improve nuclear physics models.
  - Tested emulation and calibration tools for uncertainty quantification parameters.
  - Automated interfaces between surrogate models and calibration techniques.
- 2016 – 2021 **Scientific Researcher (Junior Associate)**, *Jagiellonian University*, Kraków, Poland.
- Handled experimental nuclear physics data.
  - Conducted data analysis and visualization using various statistical models.
  - Developed numerical simulations and published scientific papers.
  - Gained knowledge in data science and machine learning.

### Teaching Experience

- 2023 **Teaching Assistant**, *AGH University of Kraków*, Kraków, Poland.
- Tutor of exercise classes in Data Analysis for students in Technical Physics of the second cycle.
- 2017 – 2019 **Teaching Assistant**, *Jagiellonian University*, Kraków, Poland.
- Led laboratory classes in Nuclear Physics for “Advanced Materials and Nanotechnologies” students every summer semester.
  - Developed statistical web applets (in Javascript) and conducted exercise classes in Probability and Statistics for students of the second year of computer science.
  - Led Physics laboratory classes for schoolchildren.

### Languages

English (Advanced), Polish (Upper-Intermediate), German (Beginner), Russian (Fluent), Ukrainian (Native)

### Courses & Certifications

- [robot\\_dreams](#), Data Science with Python
- [Data Visualization](#), [Data Analysis](#), [Data Analysis](#), [Machine Learning](#) using Python, IBM/Coursera
- [Python Programmer Track](#), [Data Scientist with Python](#), DataCamp

### Professional Interests

Statistics and machine learning, data science, calculus, numerical simulations

### Interests

- Mountaineering
- Solving mathematical problems and coding
- Mushroom hunting
- History
- Music

### Education

- Oct. 2016 – **Doctor of Philosophy in Physics**, *Jagiellonian University*, Kraków.
- Sept. 2021