

# Vangelis Kourlitis PhD

Data Science | Technical Leadership | ex-CERN

## PROFILE

Data scientist with 10 years experience in technical projects and leadership positions at CERN. Specialized in end-to-end developing scalable pipelines in big data analytics and machine learning domains. Skilled cross-functional leader with extensive experience mentoring teams and managing R&D projects. Now seeking to transition into business-oriented role, master new tooling and enhance data-driven decision-making.

## CONTACT

**Location:** Thessaloniki, Greece  
**Email:** kourlitisV@hotmail.com  
**Tel:** +30 6949552002  
**LinkedIn:** vangelis-kourlitis  
**GitHub:** ekourlit

## SKILLS

Programming

Python • C++ • CUDA • SQL • Bash

DevOps & Cloud

Git (CI/CD) • MLflow • Docker • AWS (SageMaker) • GCP

Distributed Compute

Dask • Apache Spark / Databricks • Ray

Workflow Schedulers

HTCondor • Slurm

## EDUCATION

### ATHENA RESEARCH CENTER

Athens Natural Language Processing Summer School  
2024 | Athens, GR

### UNIVERSITY OF SHEFFIELD

PhD in Physics  
2015 - 2019 | Sheffield, UK

### ARISTOTLE UNIVERSITY OF THESSALONIKI

B.Sc. in Physics  
2009 - 2015 | Thessaloniki, GR

## EXPERIENCE

### CERN & TECHNICAL UNIVERSITY OF MUNICH | Geneva, CH Analysis Model Group Coordinator

Oct 2023 - Dec 2024

- Directed a global, 120-member cross-functional team, delivering 24/7 production support and new features for a ~1 M LOC data analysis platform serving more than 3,000 users.
- Reduced organizational data storage costs threefold by adopting a lightweight columnar data format.
- Rolled out streamlined software configuration paradigm, reducing analysts onboarding time by over 95%.
- Planned and supervised 5 early-career developers projects, delivering innovative software products on time and fostering professional growth.

### Data Science Researcher

Mar 2023 - Sept 2023

- Awarded €40k grant to modernize legacy C++ data transformation tools with scalable array-based solutions, achieving 4x faster analysis and aligning with modern Pythonic data science standards.
- Piloted an LLM-based workflow to generate user-facing changelogs from software release diffs, reducing manual effort by 80%.

### ARGONNE NATIONAL LABORATORY | Chicago, US

### Data Science Researcher

Nov 2019 - Feb 2023

- Led a research team of 7 in applying advanced analysis algorithms and deep learning for complex data classification, increasing analysis coverage by 12%.
- End-to-end developed a CNN computer vision model, with multi-GPU training strategy in PyTorch, to restore fast, low-accuracy sensor images and improve simulation software cost-efficiency by up to 20%.
- Achieved 10x higher training throughput for a variational autoencoder-based anomaly detection model, applicable to fraud detection, benchmarking SambaNova RDU accelerator against NVIDIA GPUs.
- Established FAIR principles for AI models and standardized methodologies among international research teams, allowing faster review cycles and enhanced transparency across projects.
- Stress-tested Google Cloud Platform's readiness and scalability for big data analysis workflows ( $\mathcal{O}(100)$  TB), minimizing TCO by optimizing egress.

### UNIVERSITY OF SHEFFIELD | Sheffield, UK

### Doctoral Researcher

Nov 2015 - Oct 2019

- Developed and maintained C++ data-analysis software across the full development lifecycle, enabling high-throughput distributed computations on a 500k core computing grid.
- Authored the first technical report at CERN to open source a complete statistical model, setting a precedent for reinterpretation and transparency in the field.
- Taught Python programming, scientific computing and ML fundamentals to MSc students through hands-on lab sessions.