Stelios Nikolaou | Data/AI Engineer

stelios.nikolaou.cph@gmail.com | +45 50102073 | Copenhagen 2100, DK | EU Citizen
steliosnikolaou.dev | linkedin.com/in/nikolaou-stelios | github.com/stel-nik

SUMMARY

Data/AI Engineer with experience building distributed ETL pipelines for the Ministry of taxation. Background in startups with hands-on machine learning model development and deployment. Currently working at Sanos Group, processing clinical trial data in a regulated environment, translating SAS data workflows into Python. Recently built a private on-prem RAG system and LLM agent for clinical documents Q&A, with Ollama, Qdrant, FastAPI, MCP.

TECHNICAL SKILLS

- **Python**: NumPy, Pandas, PySpark, Scikit-learn, PyTorch, TensoFlow-Keras, Flask, FastAPI, OOP
- **Other Languages:** SQL, C#, R, SAS, MATLAB
- **Microsoft Azure:** Data Factory (ADF), Synapse Analytics, Data Lake, Delta Lake, Parquet
- **AWS:** VPC, Route 53, ALB, ECR, ECS Fargate, RDS, S3, IAM
- **DevOps:** Azure DevOps, Git, Docker, Kubernetes, Terraform, MLflow, CI/CD, Linux, Bash
- **AI Engineering:** RAG, LLM agents, Ollama, Qdrant, MCP, on-prem LLM deployment, prompt engineering

PROFESSIONAL EXPERIENCE

Note: Roles were fixed-term contracts unless stated otherwise.

Statistical Programmer **Mar 2026 – Apr 2026**
Sanos Group - Omicron | *Clinical Trial CRO.* Copenhagen, DK

- Performed double programming and quality control of SDTM datasets, ensuring CDISC compliance.
- Converted clinical data to SDTM/ADaM tables, following internal SOPs in a regulated clinical trial environment.
- Migrated SAS data workflows, and created a Python library to improve reproducibility for the statistics team.
- Tech: Python, SAS, R, SQL.

Data Engineer **Jan 2024 – Apr 2025**
Udviklings- og Forenklingsstyrelsen (UFST) | *Ministry of Taxation, Customs.* Copenhagen, DK

- Led Spark schema evolution for data ingestion, driving XSD/XML change analysis and vendor coordination.
- Designed 50+ ADF ETL pipelines performing mapping, cleansing, deduplication and storing data in SQL tables.
- Optimized ADF pipelines using Delta logic, orchestration and scheduling, reducing processing time by 80%.
- Implemented post-release data quality checks and monitored pipelines to ensure the consistency of datasets.
- Authored technical documentation and mentored a new team member on the event streaming platform.
- Guided sprint planning by decomposing large projects into 2-week sprints and defining technical tasks.
- Tech: Python, PySpark, SQL, Azure (Synapse, Data Factory, Data Lake), XML/JSON, Azure DevOps-Git, Jira.

Data Scientist **Jan 2022 – Jan 2023**
FaunaPhotonics | *Agro-Tech sensor insect species classification start-up.* Copenhagen, DK

- Developed deep learning models, performing iterative training, evaluation, and hyperparameter tuning.
- Conducted R&D to improve insect flight event extraction, by generating audio signal processing features.
- Programmed and refactored 2 internal Python libraries for model training and data analysis, adding new features.
- Developed REST API endpoints for 2 internal web applications (Flask, C#/.NET).
- Monitored and maintained the data infrastructure for the insect sensors in the cloud (Microsoft Azure, AWS).
- Tech: Python (Pandas, Flask, TensorFlow), SQL, C#/.NET, Azure (Data Lake, Blob Storage, ADF, Functions)

Machine Learning Engineer **Apr 2021 – Sep 2021**
Soundtracktor | *Music recommendation start-up.* Copenhagen, DK
- Developed a prototype music recommendation system that was used to support additional funding proposals.
- Extracted audio signal processing features; evaluated ML models (GMMs, CNN transfer learning, ANN search).
- Led a 4-person project team and produced technical documentation for the prototype system.
- Tech: Python (NumPy, Pandas, TensorFlow-Keras, librosa), Docker, Github-Git, Jira, Confluence.

Data Analyst **Nov 2016 – Apr 2017**
Interbalkan Environment Center (I-BEK) Thessaloniki, GR
- Conducted geospatial analysis of historical data to assess the shrinkage of Lake Koronia (Python).
- Produced a technical report comparing satellite estimates with in-situ measurements.

PROJECTS

ClinicalRAG - Private On-Prem RAG System and LLM agent. **Apr 2026**
Clinical Documents Q&A, no data leaves network - FastAPI, Qdrant, Ollama, MCP, Docker, Kubernetes.

EDUCATION

MSc in Environmental Engineering (120 ECTS) **Sept 2017 – Mar 2020**
Technical University of Denmark (DTU) Copenhagen, DK
- Thesis: Application of Machine Learning Methods for Missing Satellite Data Imputation (Grade: 12/12).
- Evaluated algorithms: EM, KNN, SVD, Robust PCA, Tucker Decomposition, PARAFAC, EWMA, ARIMA.
- Proposed a hybrid EWMA–Robust PCA method that improved imputation performance.

Diploma in Hydraulics, Soil Science and Agricultural Engineering (304 ECTS) **Sept 2007 – Nov 2015**
Aristotle University of Thessaloniki (AUTH) Thessaloniki, GR
- Linear Algebra, Multivariate Calculus, Descriptive and Inferential Statistics.
- Time Series and Multidimensional Data Analysis, Numerical Analysis and Optimization.

CERTIFICATES

Machine Learning - Stanford University **Mar 2021**
- Supervised-Unsupervised Learning, Dimensionality Reduction, Anomaly Detection, Recommender Systems.

ADDITIONAL EXPERIENCE

Music Performer and Teacher **Jan 2009 – Present**
Self Employed
- Performed in hundreds live concerts and taught music theory and performance to over 20 students.
- Arranged live shows, coordinated with musicians and venues, building collaboration and communication skills.
- Trained in classical music harmony and instruments, fostering structured thinking and independent learning.