Konstantinos Pantazis

Data Scientist | Scientific AI | Industrial Process Optimisation | Digital Twins

+30 695 5532 846 | konstantinos_pantazis@hotmail.com | Athens, Greece | Open to EU relocation
LinkedIn | Webpage | ORCID

PROFILE

Data scientist and applied mathematician with a Ph.D. in mathematics, building **scientific AI and machine-learning models for physical and industrial systems**. Experience spans **surrogate modelling, empirical/physics-informed feature engineering, optimisation, graph learning, uncertainty-aware evaluation, active learning, and production-oriented ML pipelines**. Current work focuses on a **digital food-extrusion intelligence platform** that combines recipe, process, lab-trial, and multiphysics descriptors to support virtual experimentation, calibrated prediction, and trust-aware recommendations. Strong interest in roles involving **complex systems, limited/noisy data, industrial decision support, and hybrid ML for real-world operational environments**.

TECHNICAL SKILLS

Scientific AI & Process Modelling: surrogate modelling, model calibration, automated retraining, active learning, transfer learning, uncertainty quantification, statistical inference, process optimisation.

Machine Learning & Analytics: regression/classification, clustering, feature selection, feature engineering, model evaluation, graph ML, network alignment, graph matching, recommender systems

Programming & Data Systems: Python, SQL, R, MATLAB, VSCode, Power BI, REST APIs, scikit-learn, PyTorch, TensorFlow, Polars, NetworkX.

LLMs & NLP: HuggingFace, BioBERT, LangChain, RAG, RAGAS, TruLens, Pydantic, LLM APIs, graph-based retrieval.

PROFESSIONAL EXPERIENCE

Deus Ex Machina / PIPA AI Sep. 2024 – Present
Data Scientist Athens, Greece

- Own the ML/data-science intelligence layer for a **digital food-extrusion platform**, supporting virtual recipe screening, product-property prediction, process optimisation, and guided lab experimentation.
- Integrated recipe composition, controllable operating settings, historical lab trials, and multiphysics descriptors from mechanistic modelling into reproducible ML-ready pipelines, including target-specific feature-selection workflows.
- Developed staged predictive workflows combining extrusion-success classification, target-specific regression models for final product attributes, and optimisation routines for operator-controllable process settings.
- Implemented active-learning workflows to propose informative follow-up lab trials when model confidence is low, accelerating recipe onboarding and reducing unstructured trial-and-error experimentation.
- Built end-to-end ML pipelines for heterogeneous scientific and food datasets, including feature selection, automated retraining, BioBERT-based featurisation, hyperparameter tuning, and regression/classification models for sensory and molecular-property prediction.
- Engineered and evaluated LLM/RAG and graph-based RAG workflows for biomedical and scientific information extraction using frameworks such as RAGAS and TruLens, improving extraction accuracy from below 80% to 93% .
- Communicated clear, actionable insights through reports and presentations for both technical and non-technical stakeholders.

Johns Hopkins University Jun. 2022 – Jun. 2023
Postdoctoral Fellow, Applied Mathematics & Statistics Baltimore, USA

- Developed statistical methods for network-valued and time-series-derived data, with applications to neuroscience, representation learning, dependence modelling, and downstream inference.
- Studied how network embeddings relate to classical statistical techniques such as PCA, focusing on model interpretation, dependence, effective sample size, clustering, and classification.
- Taught *Introduction to Optimization* and *Probability and Statistics for Biological Sciences and Engineering*, translating technical material for interdisciplinary audiences.

Microsoft May 2022 – Aug. 2022
Joint E+D & MSR Research Intern Seattle, USA

- Modelled security incidents as graphs and implemented deep graph-similarity learning methods, including SimGNN, for relational incident analysis and risk exploration.

- Processed network-streaming security data and built interactive Power BI dashboards for large incident collections, supporting exploratory analysis and stakeholder-facing risk quantification.

DEVCOM Army Research Laboratory Jun. 2021 – Aug. 2021
Machine Learning Intern Washington, DC, USA

- Surveyed recommender-system methods and proposed collaborative filtering approaches for prioritising information objects in dynamic, high-volume decision environments.

SELECTED TECHNICAL WORK

Scientific AI for Food Extrusion Digital Twins Industry

- Built hybrid ML workflows for a real-world extrusion platform combining empirical lab data, controllable process settings, recipe information, and digital-twin descriptors to predict product properties and recommend feasible operating conditions.

Network Inference, Graph Matching & Relational ML Ph.D., postdoc, internships

- Developed graph-matching and joint-embedding methods for multiple networks, entity alignment, connectome analysis, social platforms, security incidents, and knowledge-graph signal recovery.

EDUCATION

University of Maryland, College Park 2019 – 2022
Ph.D. in Mathematics; dissertation: Statistical Inference across Multiple Graphs USA

University of Massachusetts Amherst 2017 – 2019
Ph.D. Studies in Mathematics; transferred to UMD following advisor relocation USA

National & Kapodistrian University of Athens 2011 – 2016
B.Sc. in Applied Mathematics Greece

SELECTED PUBLICATIONS

- Hybrid multi-physics and machine-learning framework for prediction and optimization in twin-screw food extrusion. *Journal of Food Engineering*, under review, 2026.
- Optimizing the Induced Correlation in Omnibus Joint Graph Embeddings. *Journal of Computational and Graphical Statistics*, accepted, 2026.
- Clustered Graph Matching for Label Recovery and Graph Classification. *IEEE Transactions on Network Science and Engineering*, 2023.
- The Importance of Being Correlated. *Journal of Machine Learning Research*, 2022.
- Multiplex Graph Matching Matched Filters. *Applied Network Science*, 2022.