Natalia Sikora

taliaora | Natalia Sikora | natalia.sikora.jobs@gmail.com | Website

Profile

EPSRC Fellow, Applied Machine Learning Engineer and a PhD Candidate (Viva Oct'25) with experience designing and deploying scalable ML systems. Expertise across the full pipeline, from data ingestion to model deployment, using PyTorch, TensorFlow, and MLOps workflows. Published in *Nature* and AAAI-25 B1. Proven record building robust, explainable, and high-performance ML/DL models in healthcare, genomics, and multimodal data settings. Experienced in project management.

Education

Swansea University, PhD Student Sep 2020 – Apr 2025
AIMLAC CDT in AI, ML and Advanced Computing

- Developed scalable transformer and graph-based architectures for multimodal learning.
- Built unsupervised pretraining models for multi-omics integration and drug discovery.
- Published AAAI-25 B1 paper on multimodal fusion.

University of Oxford, Visiting Student Researcher, Shi Lab* Sep 2022 – Sep 2024

Main project:

- Lead researcher on a deep learning-based drug discovery project using Multi-Objective Rationale with Reinforcement Learning and Graph Neural Networks.

  Tools: HPC, cloud infrastructure, PyTorch Geometric, Nextflow, Docker, Kubernets.

University of Nottingham *BSc (Hons)* 2017 – 2020

Prfessional & Research Experience

EPSRC Postdoctoral Pathway Fellowship Oct 2025 – Feb 2026
Focusing on multiomic data integration methods. GNN, Transformers.

Freelance Data Scientist / ML Engineer Jun 2025 – Present

- Deliver ML pipelines for clients in bioinformatics and healthcare.
- Deploy trained models via Docker + Flask REST APIs; automate workflows with Nextflow.
- Applied MLOps best practices: versioning, unit tests, CI/CD, container orchestration.

The Alan Turing Institute, Data Study Group Researcher May 2024

- Improved fairness-aware financial risk model accuracy by 35% using XGBoost and AdaBoost with explainable AI and fairness methods.

DataAid, Team Leader, Synergy Essex Nov 2021 – Apr 2022

- Led and mentored a team of 12+ data scientists during the final stage hackathon; presented outcomes to the UK Ministry of Justice.
- Quantified the charity's impact, identifying a 35% increase in survivor crime reporting and a twofold rise in successful prosecutions.

- Directed a 6-month project with 4 engineers, overseeing secure data transfer, preprocessing, analysis, and comprehensive reporting on societal and financial outcomes.

DataAid, Project Lead, Fairtrade Foundation Oct 2020 – Jan 2021

- Led a 16-member data analysis team for the foundation's debut event and served as main liaison with the charity partner.
- Built a Markov Chain Monte Carlo model linking crop production to Fairtrade Premium income.
- Managed data cleaning, visualisation (Matplotlib, Seaborn, Plotly, Bokeh), and team communications.

Technical Skills

Languages: Python (NumPy, SciPy, Pandas, Keras, PyTorch, TensorFlow, scikit-learn, PEP-8), Java (basic), C++ (basic), R, SQL, Shell **ML Frameworks:** Transformers, GNNs, RL, Explainable AI (SHAP, LIME), Autoencoders, LLMs, CNNs **Dev & MLOps:** Docker, Flask/Django, Git/GitLab, Nextflow, CI/CD, API development, Kubernetes (basic), Linux **Cloud:** Azure, AWS EC2 & S3 **Practices:** Agile (Scrum, Kanban), OOP, Test-Driven Development, Code Review **Other:** HPC/Slurm.

Selected Awards & Honours

EPSRC Postdoctoral Fellowship (2025), Top 20 Young Professionals of Polish Descent (2025); Best Poster Award, Swansea PGR Conference (2024); Best Team Leader (DataAid 2022); EPSRC Fully Funded PhD Scholarship (2020–25).

Languages

Native: Polish, **Advanced**: English, **Fluent**: French, **Independent**: Italian, German, **Basic**: Hebrew.

Conference Presentations & Training

Talks:

- Multimodal Data Integration and Explainable Expert Systems, AI & ML Conference, Bristol (2025)
- Multi-Objective De Novo Drug Design for Difficult Drug Targets, AI Conference, Swansea (2023)

Posters:

- ColonScopeX: Leveraging Explainable Expert Systems with Multimodal Data for Improved Early Diagnosis of Colorectal Cancer. AAAI B-1 2025, Philadelphia, USA (2025)
- Early Detection of Colorectal Cancer Using Explainable AI, PGR Conference, Swansea (2024)

Co-organised: AI, ML & Advanced Computing Conference, Swansea (2023), 100+ attendees.
Attended: Genomics England Symposium '25 & '24, NeurIPS '23, UKRI ML & DL Conference '24, among others.

Teaching Experience

Tutor, Oxford Study Abroad Programme, Oriel College, University of Oxford 2022 – 2023

Swansea University 2020 – 2024
Delivered lectures, labs, and workshops in Medicine, Computer Science, and Physics.
Courses taught include: Big Data and ML (Y3/MA), AI (Y3/MA), ML (MA), Computational Probability

(Y1), Essential Skills for Physicists (Y1), Cryptography and IT-Security (Y3/MA), Mathematics for Physicists II, Bioinformatics.

Selected Publications

1. Hosseini A, Dhall A, Ikonen N, Sikora N, Nguyen S, Shen Y, et al. Perturbing LSD1 and WNT rewires transcription to synergistically induce AML differentiation. Nature. 2025 Apr 16;

2. Sikora N, Manschke RL, Tang AM, Dunstan P, Harris DA, Yang S. ColonScopeX: Leveraging Explainable Expert Systems with Multimodal Data for Improved Early Diagnosis of Colorectal Cancer. 2025 Apr 9 [cited 2025 Apr 30]; Available from: https://arxiv.org/pdf/2504.08824

3. Ayvaz DS, Belenguer L, He H, Kanubala DD, Li M, Low S, et al. Measuring Fairness in Financial Transaction Machine Learning Models. 2025 Jan 18;

Under Review:

First author

- *A Trajectory-Aware Graph Transformer for Rare-Cell Discovery in Single-Cell Multi-Omics.*
- *Explainable Deep Learning and Transfer Learning for Enhancer-Targeting Therapeutics in Ovarian Cancer.*
- *Optimising single cell multiomic integration methods and its impact on personalised medicine applications.*
- *Towards Precision Oncology: Linking Enhancer-Driven Networks to Druggable Protein Complexes.*

Co-author

- *Twelve Angry Models: Multi-Agent Plenary Discussion For Bias Awareness and Reduction.*
- *From Bots to Belief: Modelling the Spread of Automated Agendas in Online Ecosystems.*
- *Distinct H3K27 methylation states drive cellular responses to the histone demethylase inhibitor GSK-J4 in ovarian cancer cells.*