Yulia Volkova

Senior Data Scientist

📍 London, UK | LinkedIn | +447450514095 | yuulia.volkova@gmail.com

Senior Data Scientist with 8 years of experience delivering scalable ML solutions and NLP systems in production. Expertise in LLM evaluation & post-training, statistical modelling, and MLOps/infra. Passionate about AI alignment and technical safety, with a focus on mechanistic interpretability and AI control.

AI Alignment & Technical Safety

Visiting Member at London Initiative for Safe AI (LISA)   **Sept 2025 - present**

SPAR Fellowship   **Oct 2025 - present**

- CoT monitorabilty project under Qiyao Wei (MATS)

Algoverse AI Safety Research Fellowship   **Jul–Aug 2025**

- Two concurrent research projects:
    1. Mechanistic interpretability under Nicky P
    2. Cooperative AI under Max Kleiman-Weiner

ARBOx2 Oxford AI Safety Bootcamp   **Jul 2025**

- Participated and completed the ARENA curriculum.

Work Experience – Behavox

Behavox is a security software company specialising in communication surveillance. Leading AI compliance solution in the marketplace.

Data Science Production Team Lead   **2024–2025**

- Leading development and deployment of multilingual NLP applications
- Executing cross-team roadmapping for ML features under real-world production constraints
- Building reusable tooling enabling rapid MVPs for client deployments and proofs of concept

Insider Crime Product Lead   **2020–2024**

- Led R&D and production of high-recall (81% at 0.5% FPR) misconduct detection systems using RoBERTa + SBERT models architecture pipeline
- Specialised in working with Large Language Models (LLM), fine-tuning them to meet Insider Crime product requirements

- Validated and proved that shorter text sequences in RoBERTa training maintain performance, doubling training speed and halving company training costs
- Developed and maintained cloud-based infrastructures using AWS, Docker + Slack API
- Maintained model performance through continuous training, synthetic data QA, and AWS S3 workflows
- Developed a hybrid recommendation system leveraging RoBERTa and SBERT for text embeddings, K-Means for clustering, and integrated user attributes into a CatBoost model for personalised predictions

Data Scientist   **2017–2020**

Key responsibilities: development of multilingual applications and reports to flag statistical deviations and compliance breaches in client data, designing and building infrastructure to support application creation, QA and maintenance

- Landed proof of concepts that resulted in 3+ years client contracts
- Improved compliance model recalls from 70% to 90%
- Reduced client bug requests from 1/day to 1/month by setting up an efficient QA process
- Self-studied Java to remove the Analytics dependency on Backend teams
- Developed best practices for code quality as one of the first Data Science hires and created foundational documentation during company scaling

Education

MSc Economics, London School of Economics   2017–2018
— Distinction (dissertation); Advanced Macroeconomics · Econometrics · Quantitative Methods

BSc Philosophy & Economics, London School of Economics   2013–2016
— Distinction; Mathematical Methods · Statistics · Econometrics · Philosophy of Science

Skills Summary

Category	Skills
Languages & Tools	Python, Pandas, NumPy, Scikit-learn, FastAPI, spaCy, SQL, Java, Kotlin, Groovy, Jenkins
Machine Learning & NLP	LLM fine-tuning (LoRA/qLoRA), PyTorch, HuggingFace, prompt engineering, transformers, SBERT, PCA, t-SNE, models evaluation (precision, recall, ROC/PR-AUC, F1), Weights & Biases
Infrastructure	Testing automation, synthetic data generation, containerised deployment, AWS S3, GCP, Docker, Nomad, Kubernetes, TeamCity, Gitlab CI/CD pipelines
Statistics & Econometrics	Hypothesis testing, time-series analysis, A/B testing, IV estimation, panel regression, fixed/random effects, vector auto-regressions

Category	Skills
Leadership	Mentorship, technical interviewing, product roadmapping, cross-team communication, internal presentations