

Ruman Shaikh

London, UK · s.k.ruman.0@gmail.com · +44 7405 510424 · [LinkedIn](#) · [GitHub](#) · [Website](#)

EXPERIENCE

IBM — AI Engineer | *Bangalore, India · Oct 2023 – July 2025*

AI & Hybrid Cloud, WatsonX Client Engineering

- ▶ **Multi-Agent AI Systems:** Designed and deployed production multi-agent orchestration frameworks using LangGraph state machines and CrewAI; exposed as MCP servers via FastAPI with async processing and LangSmith observability; reusable enterprise-grade patterns adopted across multiple client engagements.
- ▶ **NL-to-SQL:** Built natural-language-to-SQL interfaces enabling non-technical business users to query complex relational databases accurately; removing analyst bottlenecks and accelerating self-service data access.
- ▶ **LLM Fine-Tuning & Benchmarking:** Fine-tuned LLMs for specialised enterprise domains using LoRA-based SFT on real and synthetic datasets; built reproducible benchmarking pipelines (zero-shot, few-shot, fine-tuned) with standardised metrics (F1, ROUGE, BLEU). Applied layer pruning and knowledge distillation for SLM deployment.
- ▶ **Computer Vision & Time-Series:** Built end-to-end CV pipelines on satellite imagery (preprocessing, training, deployment) for climate risk modelling; developed time-series forecasting models on large-scale environmental datasets for enterprise risk applications.
- ▶ **RAG Architecture:** Engineered high-performance RAG systems with hybrid vector retrieval; enterprise form-filling assistant processing thousands of documents achieved ~80% reduction in end-to-end workflow time.
- ▶ **MLOps & Data Infrastructure:** Built end-to-end MLOps pipelines for model monitoring and lifecycle management achieving ~90% drift-detection accuracy (k-fold cross-validated on stratified subsets); architecture designed to scale across petabyte-scale distributed data lakes (PySpark / Hadoop ecosystems).
- ▶ **Content Standardisation at Scale:** Delivered parallel LLM execution pipeline for large-scale document standardisation, significantly reducing compliance violations and manual review overhead across client workflows.

Oracle Health (formerly Cerner) — Software Engineer | *Bangalore, India · June 2021 – Aug 2023*

Population Health — Software & Data Engineering

- ▶ **CMS-Compliant Data Pipelines:** Designed and maintained population health pipelines adhering to CMS guidelines, ingesting EHR, and lab data across millions of patients served by hundreds of US healthcare systems nationwide; produced standardised population health reports for hospital and payer clients.
- ▶ **Population Health Analytics:** Built analytics workflows enabling providers to track patient cohort trends, chronic disease prevalence, and care gaps, feeding downstream dashboards and CMS quality reports.
- ▶ **Test Engineering:** Built a Python test automation pipeline covering unit, integration, and black-box testing across the population health platform, enforced rigorously given CMS compliance requirements on patient data; reduced regression risk across every pipeline release.
- ▶ **Performance Engineering:** Achieved 90% reduction in pipeline processing time (days → under one day) through query optimisation, partitioning strategies, and distributed restructuring.
- ▶ **ETL Automation:** Built automated client data workflows using Apache Airflow and PySpark, improving report generation reliability and reducing manual intervention in production environments.
- ▶ **Software Delivery:** Delivered production features using Python, Oracle SQL, Java, and Git across agile sprint cycles tracked in Jira.

EDUCATION

Imperial College London — MSc Biomedical Engineering | *London, UK · Sept 2025 – Sept 2026*

Specialisation: Computational Bioengineering · Supervisor: Prof. Reiko Tanaka

- ▶ **Dissertation:** Physics-Informed Operator Learning for Microbial Dynamics - PI-DION and QCPINN architectures for generalised inverse modelling of biological dynamical systems (generalised Lotka-Volterra) in data-scarce settings, supervised by Prof. Reiko Tanaka.
- ▶ **Electives:** Brain Machine Interface, AI for Drug Discovery, Biomimetics
- ▶ **Tools:** PyTorch, PennyLane, Optuna, Weights & Biases, Ray, Hydra, Docker

New Horizon College of Engineering — BEng Computer Science & Engineering | *Bangalore, India · July 2017 – Aug 2021*

- ▶ CGPA: 9.09 / 10 (First Class)
- ▶ Relevant coursework: Data Structures & Algorithms, Operating Systems, Database Management Systems, Computer Networks, Machine Learning, Software Engineering.

PROJECTS

Impact Scholar — Architectural Priors in Biological Neural Agents | *NeuroMatch Academy – Sept 2024 – Apr 2025*

[PyTorch](#) · [model compression \(distillation, pruning, quantization\)](#) · [connectomics](#) · [scikit-learn](#)

- ▶ Investigated whether structural priors derived from biological neural circuits (C. elegans locomotion, quadruped running connectomes) outperform equivalent-parameter MLPs on motor control tasks.
- ▶ Applied model compression techniques: distillation, pruning, quantization, and systematic parameter grid search to identify minimal model configurations matching biological performance.
- ▶ Demonstrated that biologically evolved architectures prior informed networks outperform standard MLP scaling laws in terms parameter efficiency. Published finding in Zenodo (2025).

Autonomous Ice Hockey Agent | *UT Austin (online) – May 2023*

[PyTorch](#) · [Ray](#) · [OpenAI Gym](#) · [REINFORCE](#) · [distributed training](#)

- ▶ Built image-based and state-based RL agents to play ice hockey, comparing both observation modalities across sample efficiency and generalisation.
- ▶ Designed a full RL pipeline from scratch: reward function engineering, environment rollout design, data collection, training loop, and policy-gradient optimisation using REINFORCE algorithm.
- ▶ Achieved 85% ball-tracking accuracy and >80% overall game win rate through iterative reward shaping and hyperparameter optimisation.
- ▶ Parallelised environment rollouts using Ray for distributed execution, significantly reducing convergence time and improving training throughput.

Prediction of Motion States from ECoG Recordings | *NeuroMatch Academy (Summer School) – July 2023*

[scikit-learn](#) · [pandas](#) · [matplotlib](#) · [SciPy](#) · [signal processing](#) · [PCA](#)

- ▶ Built end-to-end neural signal pipeline correlating ECoG brain recordings with continuous 2D cursor movement trajectories, targeting real-time BCI decoding.
- ▶ Applied frequency-domain filtering and PCA-based dimensionality reduction to isolate informative neural components and reduce processing latency to 5 ms.
- ▶ Achieved max R^2 of 49.3% for continuous cursor velocity prediction; identified statistically significant correlations between specific Brodmann cortical areas and directional movement signals, enabling spatially targeted feature extraction.

Analysing Dataset Artifacts Using ELECTRA | *UT Austin (online) – June 2023*

[PyTorch](#) · [HuggingFace Transformers](#) · [NLP](#) · [fine-tuning](#) · [NLI](#)

- ▶ Fine-tuned ELECTRA for natural language inference (NLI) achieving 88.24% accuracy on multi-class entailment classification.
- ▶ Designed systematic error analysis framework to identify and categorise dataset artifacts spurious correlations in training data that inflate accuracy without genuine language understanding.
- ▶ Demonstrated artifact correction significantly improves out-of-distribution robustness; extended fine-tuning across diverse datasets to validate generalisation beyond training distribution.

PUBLICATIONS

[Muratore, P., Moradimanesh, Z., Shaikh, R. A., Zargarnezhad, N., Meek, A., & Kuzovkin, I. \(2025\). Biological Connectivity Patterns as a Blueprint for Efficient Neural Architectures in Reinforcement Learning \(v1.0\). Zenodo.](#)

Shaikh, R. A., Hazra, S., Nambiar, N. S. B., & Uma, N. (2021). Intelligent Block Energy. International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET).

TECHNICAL SKILLS

Languages & Frameworks: Python, Java, C, C++, SQL, HTML/CSS, React.js, PyTorch, TensorFlow, scikit-learn, OpenCV, LangChain, LlamaIndex, CrewAI, NLTK, HuggingFace Transformers, Streamlit

AI / ML: LLM fine-tuning (LoRA/SFT), RAG, retrieval-augmented generation, multi-agent orchestration, MCP servers, LangGraph, NL-to-SQL, computer vision, object detection, time-series forecasting, NLP, reinforcement learning, model compression, prompt engineering, vector databases

Data & MLOps: PySpark, Hadoop, Hive, Apache Airflow, Oracle SQL, MLOps, model monitoring, model drift detection, Docker, IBM Code Engine, Git, GitHub, LangSmith, Weights & Biases, Optuna

Biomedical / Signal: ECoG signal decoding, BCI, PINNs, physics-informed neural networks, operator learning (DeepONet / PI-DION), CUDA, GPU programming, computational neuroscience, ODE systems, population health informatics (CMS)

Platforms & Cloud: IBM WatsonX (watsonx.ai, watsonx.data, watsonx.governance), IBM Code Engine, FastAPI, Apache Spark, Keras, statsmodels, SciPy, NumPy