Shengyuan Liu

liushengyuan2006@outlook.com | linkedin.com/in/liu-shengyuan | +1 (201) 204 2783

Education

New York University — New York, US

Master of Science in Data Science — Sep 2026 – May 2028

- **Relevant Courses**: Introduction to Data Science, Machine Learning, Programming for Data Science, Data Engineering, Advanced Python.

Imperial College London — London, UK

Bachelor of Science in Mathematics, First Class Honours — Sep 2023 – Jun 2026

- **Relevant Courses**: Data Science, Statistical Learning, Optimisation, Time Series, Computational Linear Algebra, Game Theory, Stochastic Process, SDE in Financial Modelling.

Professional Experience

Institute of Automation, Chinese Academy of Sciences — Oct 2025 – Dec 2025

Research Assistant

- Reviewed video multimodal reasoning and **GRPO** variants to study adaptive rewards. Ran original TinyLLaVA-Video-R1 training and benchmark evaluations following the paper's setup.
- Prototyped GRPO-LEAD-inspired modifications using within-group answer accuracy to estimate question difficulty, **difficulty-aware length rewards** and **advantage reweighting**, then initiated modified-model training.

Neusoft Group — Jul 2025 – Sep 2025

Software Engineer

- Processed 10K+ CheXpert Plus frontal X-rays with complete findings and structured reports. Adapted **MedGemma-4B** for Chinese reporting generation using **4-bit quantization, LoRA** and two A100 GPUs.
- Built a Deepseek **evaluation pipeline** measuring disease agreement, omissions, false positives and descriptive accuracy. Fine-tuned scores rose from **40+ to 80/100** against baseline.
- Built an OCR, de-identification and regex extraction pipeline to structure reports from 2K private mammography cases, enabling MedGemma fine-tuning for **BI-RADS prediction**.

Projects

KnowledgeBase — Sep 2026 – Present

Personal Project: RAG, FastAPI, React, Chroma, SQLite

- Built a citation-grounded RAG pipeline with **Docling parsing and structural chunking**, preserving source spans, table headers, and page references; enabled original-page citation navigation and retained historical evidence after source deletion.
- Integrated **BM25, dense retrieval, RRF, BGE reranking, and parent–child chunking** with LLM query decomposition and multihop retrieval, improved annotated evidence character coverage from **90.58% to 99.96%** and full-span recall from **59.38% to 68.75%** across 16 questions within a **4,096-token evidence budget**.
- Built reproducible evaluations for retrieval, citations, abstention, latency, and token usage, matched PDF regression passes **6/6**, with **13/13 expected facts supported and cited**.

Kaggle Pokemon TCG AI Battle — Jun 2026 – Aug 2026

Silver Medal, 52 / 6,807 teams

- Built a PyTorch agent via behavior cloning of public game-play and **PPO**. Early PPO beat its cloning parent in 52.8% of 6,000 seat-balanced local games.
- Designed **state and action encoders** with shared card embeddings, a shared Pokemon object encoder and permutation-invariant pooling for variable legal actions. Distilled 10 matchup specialists via cross-entropy into an 8M-parameter policy.
- Profiled CPU inference and benchmarked **ONNX** Runtime against PyTorch, reducing model-forward latency by **44%**, from **3.45 ms to 1.93 ms**, in single-threaded local tests.

Forecast Combination Puzzle — May 2025 – Jun 2025

Research Project: Time Series Analysis, Random Forest, Boosting

- Implemented exponential smoothing, Holt–Winters, ARIMA, Random Forest and Boosting models in Python on an M4 daily time series; evaluated forecast accuracy using MSE, RMSE, and MAE.
- Compared equal-weight, error-based, and covariance-based forecast combinations, found simple averaging outperformed error-based weighting and analyzed how residual correlations affected combination performance.

Technical Skills

Programming: Python (PyTorch, NumPy, Pandas, Matplotlib, FastAPI), TypeScript/JavaScript, HTML & CSS, Go

Tools: PostgreSQL, Git, Linux, Shell, Codex, Claude Code.