Haoyan Luo

Homepage: https://clarenceluo78.github.io/

Email : h.luo23@imperial.ac.uk
Mobile : +44 7707-977-358

Education

Imperial College London — London, UK
M.Res. in AI & Machine Learning, Full Scholarship; Distinction — *Sep. 2023 - Sep. 2024*

- **Research topics**: Explainable and Controllable Language Models, Interpretability and its Applications
- **Thesis**: Explainability in Large Language Models: Pathways to Refinement and Alignment
- **Supervisor**: Prof. Lucia Specia

The Chinese University of Hong Kong, Shenzhen — Shenzhen, China
B.Eng. in Computer Science; First-Class Honours — *Sep. 2019 - July 2023*

- **Honors**: Dean's List of School of Data Science (AY21-23), Undergraduate Research Awards (AY21-23), Undergraduate Student Teaching Fellow Award (AY21-22), Shaw College Outstanding Graduates Award (Top 1%)
- **Services**: President of TEDxCUHK(SZ), Coordinator of Shaw College Student Union, Host Coordinator of CPRO

Publications

1. **Haoyan Luo**, Lucia Specia. **Bias in the Distributed Subspaces: A Mechanistic Debiasing Approach via Localized Subspace Probing**, *Under Review*.
2. **Haoyan Luo**, Lucia Specia. **Tuning Language Models by Mixture-of-Depths Ensemble**, *Under Review*.
3. **Haoyan Luo**, Lucia Specia. **From Understanding to Utilization: A Survey on Explainability for Large Language Models**, *arXiv*, 2024.
4. Tianping Zhang, Zheyu Zhang, Zhiyuan Fan, **Haoyan Luo**, Fengyuan Liu, Wei Cao, Jian Li. **OpenFE: Automated Feature Generation with Expert-level Performance**, *Fortieth International Conference on Machine Learning (ICML)*, 2023.
5. **Haoyan Luo**, Xiaofan Gui, Wei Cao, Jiang Bian. **ActiveAD: Enhancing Anomaly Detection in Tabular Data through Active Learning Strategies**, *arXiv*, 2023.

Research & Work Experience

- **J.P. Morgan** — London, UK
  AI & Data Science Associate — *Sep. 2024 - Present*
  - **TIDE - Topic Insigts and Document Exploration** : Leading the development of topic modeling, synthetic data generation, and topic distillation pipelines in TIDE, our state-of-the-art solution powered by LLMs.
  - **LLM Agents Hallucination Detection** : Developing a unified hallucination detection pipeline tailored for use-case-driven agents.

- **Shanghai AI Laboratory** — Shanghai, China
  Research Intern, supervised by Dr. Xun Zhao & Dr. Dahua Lin — *July 2023 - Jan. 2024*
  - **Mechanistic LLM Interpretation**: Utilize the trained or customized LM head, discover that the hidden states of transformers can be viewed as large memory blocks, with each of its cells being highly decomposable and explainable;
  - **Inference-time Intervention**: Leverage insights from interpretation to identify specific components or 'pipelines' within the models that contribute to undesirable behavior. Develop methods that enable targeted, low-cost interventions to control model behavior;

- **Microsoft Research Asia** — Beijing, China
  Research Intern, supervised by Dr. Jiang Bian & Dr. Wei Cao — *July 2022 – Nov. 2022*
  - **Feature engineering**: Explore various automated feature generation methods, and co-proposed OpenFE, a novel framework with feature boosting method to accurately identify useful new features for tabular data. Open-sourced our code with 500+ stars;
  - **Active anomaly detection**: Research on autoencoder-based anomaly detection methods. Survey existing active learning methods in various domains. Conduct extensive experiments combining AL methods into anomaly detection problems. Proposed a meta-learning labelling framework in the context of active anomaly detection (code);

- **Deep Reinforcement Learning Research Team, CUHK(SZ)** — Shenzhen, China
  Student Researcher, supervised by Prof. Shuang Li — *May 2022 – Nov. 2022*
  - **Meta-policy on temporal point process**: Research multivariate temporal point process and formulate the traffic congestion and epidemic (COVID-19) curbing problems as finite-time horizon model-based RL problems and apply Neural ODEs to model events over space and time. Embed temporal point process dynamics into a meta-policy learning framework using graph neural networks and extract explainable and transferrable information from the learned trajectories (code);

- **Tencent** Shenzhen, China
  Big Data Engineer Intern *June 2021 – Sep. 2021*
  - **Recommendation system**: Built a recommendation system sub-module with a session-based recurrent neural network using PyTorch and deployed it within the "news" page and "Daily Q&A" page in company's WeChat Mini Program;
  - **Data warehouse construction**: Added Hive and ClickHouse support and used Scala (together with SQL insertion and Spark & Hadoop embedded features) to query useful information from different layers in the data warehouse for further data analysis;

- **Shenzhen Research Institute of Big Data** Shenzhen, China
  Research assistant, supervised by Prof. Xiaodong Luo *Oct. 2020 – March 2021*
  - **Sparse linear programming**: Researched and deployed HiGHs software for solving large-scale sparse linear programming (LP) and mixed-integer programming (MIP) in a real-world logistic dataset, Assembled algorithms and used Doxygen to generate documentation from annotated MATLAB and C++ sources;

- **Seasun Entertainment, Kingsoft** Zhuhai, China
  Software Engineer Intern *June 2020 – August 2020*
  - **Backend development**: Built and maintained a large-scale intranet used to monitor users' behaviors and dialogue contents in a popular web game in China, JX3, with over 5 million active users;

Selected Projects

- **Converting to Realistic Professional Singing Voices with Singer-Adaptive Representations**: Finding the probable weakness of the existing SVC systems in converting to professional singing voice and building a model with generalizability across various target vocalists while preserving high-quality voice conversion (code).
- **Steering the Networked Temporal Point Processes via Controlling the Network Graph**: Developed a model-based reinforcement learning strategy for optimizing network dynamics, using neural ODEs to model multivariate temporal point processes. Created a novel approach that manipulates graph topology for effective intervention in networks, such as mitigating epidemics or easing traffic congestion.(code).
- **Active Learning for Anomaly Detection on Tabular Data**: Propose a pipeline, ActiveAD, for active anomaly detection that combines anomaly detection models and active learning querying strategies to improve the efficiency and effectiveness of identifying anomalies with limited labeled data (code).
- **Parallel N-body and Heat Simulation**: Implemented parallel computing programs in MPI, Pthread, OpenMP, CUDA, and MPI + OpenMP hybrid methods in C++, monitored and analyzed performance through comprehensive experiments on school's HPC (code).
- **Deep-Learning-Lookup: Deep Learning Algorithms and Utils in Python**: Implemented various deep learning code examples (notebooks and python scripts) and useful utility functions including data manipulation, visualization, training, and evaluation (code).
- **Mindy: A Corporate Management and Mind Mapping Application**: Incorporated and implemented mind map web application with VUE and Django. Designed a database with high availability and scalability using MySQL to facilitate our web application (code).
- **MIPS Pipe-lined CPU Hardware Design**: Designed and implemented a five-stage-pipeline MIPS processor in Verilog. Solved control and data hazards by stalling, forwarding and implementing auxiliary MIPS ISA assembler and simulator (code).

Competitions and Awards

- Endowment Full Scholarship, Imperial College London 2023
- Meritorious Winner, Mathematical Contest in Modeling (MCM'22) 2022
- First Prize in 13^{th} National Mathematical Competition for College Students, China 2021
- First Prize in Contemporary Undergraduate Mathematical Contest in Modeling, China (CUMCM'21) 2021
- Third Prize, RoboMaster University AI Challenge 2021
- National Silver Award, First Prize Award in Guangdong, Google App Inventor Challenge 2019

Technical Skills

- **Programming Languages**: Python, C++, Java, JavaScript, Node.js, MATLAB, Scala
- **Tools**: Git, Vim, LaTeX, Docker, (DB) MySQL, MongoDB, Click House, (Machine Learning) PyTorch, Scikit-learn, (Web) VUE, Django, Spring Boot
- **Platforms**: Linux, (Robotics) ROS platform, (Big Data) Apache Spark
- **Languages**: Mandarin Chinese (Native), Cantonese (Fluent), English (Fluent, GRE: 329, TOEFL: 110)