Cheryl LEE (Baitong)

https://cheryllee.vip | cheryllee@link.cuhk.edu.hk | cherylleecuhk

Education

- **The Chinese University of Hong Kong** — Hong Kong SAR
  Ph.D. - Computer Science and Engineering; Supervisor: Prof. Michael R. Lyu — *Aug 2022 – Present*
  - **Research Interest**: Developing deep learning-based / optimization models to facilitate the reliability of cloud-scale software systems (AIOps).
  - **Highlighted Courses**: Foundations of Optimization, Graph Mining, Natural Language Processing, Data Science in Economics.
- **Cornell University** — NYC, U.S.
  Master - Operation Research & Information Engineering; GPA: 3.9/4.0 (Rank Top 1%) — *Aug 2021 – June 2022*
  - **Highlighted Courses**: Applied Machine Learning, Deep Learning, Optimization Methods, Modeling Under Uncertainty, E-Logistics.
- **Peking University** — Beijing, China
  Bachelor - Computer Science & Technology; GPA: 3.5/4.0 (Rank Top 30%) — *Aug 2016 – June 2020*
  - **Highlighted Courses**: Probability Theory, Statistics, Game Theory, Applications of Big Data Techniques, Database Systems, Data Structure and Algorithm, C++ Programming.

Experience

- **The Chinese University of Hong Kong** — Hong Kong SAR
  Research Assistant — *Jul 2021 - Jul 2022*
  - **Cross-modal Anomaly Detection**: Proposed a cross-modal attention-based approach to fuse text-based logs and multivariate metric time series for heterogeneous anomaly detection. Published a conference paper in ICSE'2023.
  - **Microservice Root Cause Localization**: Proposed a multi-modal GNN-based approach for troubleshooting microservices, integrating anomaly detection and root cause localization into an end-to-end framework. Published a conference paper in ICSE'2023.
- **Apple Inc.** — Beijing, China
  Machine Learning Engineer (Internship) — *Jul 2020 - Dec 2020*
  - **Log Anomaly Detection**: Proposed an unsupervised "Gradual Clustering" log parser to analyze watch production logs; Designed a Transformer-based detector and achieved over 98.29% top-1 accuracy.
  - **Duplicated Issue Identification**: Proposed a document-level algorithm based on Longest Common Substring to identify duplicated issues.
- **Deloitte Touche Tohmatsu CPA Ltd.** — Beijing, China
  Risk Analyst (Internship) — *Jul 2019 - Sep 2019*
  - **Report Analysis Automatization**: Leveraged BiLSTM-CRF to recognize name entities and extract relationships from the financial reports; Aligned the entities via BIRCH clustering.
  - **Company Appraisal**: Devised an appraisal system to quantify the operation and potential of electric companies.
- **Baidu Inc.** — Beijing, China
  Data Analyst (Internship) — *Jul 2018 - Sep 2018*
  - **User Demand Incubation**: Incubated a traffic routing function by monitoring and mining behavior logs of users interacting with a voice-assisted smart device. The function's Page View achieved the top 5.
  - **Market Performance Investigation**: Investigated and reported the market performance of the low-price selling strategy of similar devices.

Skills

- **Coding**: Python (proficient), C++, SQL — **Frameworks**: PyTorch, Pandas, Spark
- **Languages**: English, Mandarin

Activities

- **Teaching Assistant**: (CSCI3100) Software Engineering, (CSCI2720) Building Web Applications — 2022-2023
- **Ethnic Minority (Manchu) Culture Communication Club**: Founder — 2018-2020

Publications

ICSE 2023 — **C. Lee**, T. Yang, Z. Chen, Y. Su, Y. Yang, M. R. Lyu, "Heterogeneous anomaly detection for software systems via semi-supervised cross-modal attention," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, 2023, pp. 1724–1736, DOI: 10.1109/ICSE48619.2023.00148 .

ICSE 2023 — **C. Lee**, T. Yang, Z. Chen, Y. Su, M. R. Lyu, "Eadro: An End-to-End Troubleshooting Framework for Microservices on Multi-source Data," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, 2023, pp. 1750–1762, DOI: 10.1109/ICSE48619.2023.00150 .

ASE 2023 — **C. Lee**, T. Yang, Z. Chen, Y. Su, Y. Yang, M. R. Lyu, "Maat: Performance Metric Anomaly Anticipation for Cloud Services with Conditional Diffusion," in *Proceedings of IEEE/ACM 38th International Conference on Automated Software Engineering (ASE)*, 2023, URL: arxiv.org/abs/2308.07676 .

ICSE 2023 — Y. Huo, Y. Su, **C. Lee**, M. R. Lyu, "SemParser: A Semantic Parser for Log Analytics," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, 2023, pp. 881–893, DOI: 10.1109/ICSE48619.2023.00082 .

ISSRE 2023 — Y. Huo, **C. Lee**, Y. Su, S. Shan, J. Liu, M. R. Lyu, "EvLog: Identifying Anomalous Logs over Software Evolution," in *Companion Proceedings of IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*, 2023, URL: arxiv.org/abs/2306.01509 .

ISSREW 2022 — T. Yang, **C. Lee**, J. Shen, Y. Su, Y. Yang, Y. Yang, M. R. Lyu, "Managing Service Dependency for Cloud Reliability: The Industrial Practice," in *Companion Proceedings of IEEE 33rd International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2022, pp. 67–68, DOI: 10.1109/ISSREW55968.2022.00041 .

Arxiv 2022 — T. Yang, **C. Lee**, J. Shen, Y. Su, Y. Yang, Y. Yang, M. R. Lyu, "An Adaptive Resilience Testing Framework for Microservice Systems," in *arXiv preprint*, 2022, URL: arxiv.org/abs/2212.12850 .

SCI Q1 2022 — F. Liu, Y. Li, **B. Li**, "A novel probabilistic framework with interpretability for generator coherency identification," *International Journal of Electrical Power & Energy Systems*, vol. 143, p.108474, 2022, ISSN:0142-0615.

SCI Q1 2021 — F. Liu, Y. Li, **B. Li**, J. Li, H. Xie, "Bitcoin transaction strategy construction based on deep reinforcement learning," *Applied Soft Computing*, vol. 113, p. 107952, ISSN:1568-4946.

Patents

1. M. Lyu, **B. Li**, T. Yang, Z. Chen, and Y. Su, "A microservice fault diagnosis method and system," CN202211368449.4, 2022.
2. Z. Yang, F. Liu, and **B. Li**, "A faster-than-real-time observation and analysis method of voltage quality based on fpga," 2 022 109 279 359, 2022.
3. **B. Li**, F. Liu, H. Xie, W. Qi, and T. Yuan, "Big data analysis platform for ac / dc power grid with a high proportion of alternative energy," CN2020SR1652151, 2020.
4. Q. Shi, **B. Li**, F. Liu, H. Xie, and J. Zhai, "The platform for the intelligent identification of coherency generator clusters in the power system," CN2020SR1740513, 2020.

Honors and Awards

- **Honors**
  - **Postgraduate Studentship**, The Chinese University of Hong Kong *2022*
  - **Merit Scholarship**, Cornell University *2020*
- **Competition Awards**
  - **National 3rd Prize**, The 17th Challenge Cup *2021*
  - **First Prize**, The 4th Baidu star entrepreneurship competition *2020*
  - **Meritorious Winner**, Mathematical Contest in Modeling *2018*