Cheryl (Baitong) LEE

🌐 https://cheryllee.vip | ✉ cheryllee@link.cuhk.edu.hk 🎓 Google Scholar

EDUCATION

The Chinese University of Hong Kong   Hong Kong SAR
Ph.D. - Computer Science and Engineering; Advisor: Prof. Michael R. Lyu   *Aug 2022 – Present*

- **Research Interest**: AI for software engineering automation, *incl.* fault diagnosis, debugging, resource optimization.
- **Courses**: Foundations of Optimization, Natural Language Processing, Pattern Recognition, Data Science in Economics.

Cornell University (Cornell Tech)   NYC, U.S.
Master - Operation Research & Information Engineering   *Aug 2020 – June 2021*

- **Courses**: Machine Learning, Deep Learning, Optimization Methods, Modeling Under Uncertainty, E-Logistics.

Peking University   Beijing, China
Bachelor - Computer Science & Technology   *Aug 2016 – June 2020*

- **Courses**: Probability Theory, Statistics, Game Theory, Set Theory and Graph Theory, Mathematical Logic Big Data Techniques, Database Systems, Data Structure, Design and Analysis of Algorithms, Programming Design (C++).

RECENT PUBLICATIONS

CCF-A
2025

C. Lee, S. C. Xia, J. Huang, S. Li, L. Zhang, M. R. Lyu, "Rubber Duck, Listen to My Code: Unified Automated Debugging via LLM Agent Synergy," in *2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)*, Under Review

CCF-A
2024

C. Lee, T. Yang, Z. Zhu, Y. Huo, Y. Su, P. He, M. R. Lyu, "SPES: Towards Optimizing Performance-Resource Trade-Off for Serverless Functions," the 40th IEEE International Conference on Data Engineering (ICDE), 2024, 🔗 DOI: arxiv.org/abs/2403.17574

CCF-A
2024

Z. Zhu, **C. Lee**, X. Tang, P. He, "HeMiRCA: Fine-grained root cause analysis for microservices with heterogeneous data sources," ACM Transactions on Software Engineering and Methodology (TOSEM), Just Accepted.

CCF-A
2024

T. Yang, **C. Lee**, J. Shen, Y. Su, Y. Yang, Y. Yang, M. R. Lyu, "An Adaptive Resilience Testing Framework for Microservice Systems," in *the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA) 2024* 🔗 URL: arxiv.org/abs/2212.12850

CCF-A
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

CCF-A
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

CCF-A
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, 🔗 DOI: 10.1109/ASE56229.2023.00082

CCF-A
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

CCF-B
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, 🔗 DOI: 10.1109/ISSRE59848.2023.00018

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.
  - **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.

- **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 longest common substring-based algorithm 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 Investigation**: Investigated the market performance of the low-price selling strategy of similar devices.

HONORS AND AWARDS

- **Postgraduate Studentship** - The Chinese University of Hong Kong — *2022*
- **Merit Scholarship** - Cornell University — *2020*
- **National 3rd Prize** - The 17th Challenge Cup — *2021*
- **Bronze Prize** - Internet Plus College Students' Innovation and Entrepreneurship Competition — *2021*
- **Meritorious Winner** - Mathematical Contest in Modeling — *2018*
- **Provincial 1st Prize** - Chinese Physics Olympiad — *2015*
- **Provincial 2nd Prize** - Chinese Mathematical Olympiad — *2015*

PATENTS

- M. Lyu, **B. Li**, T. Yang, Z. Chen, and Y. Su, "A microservice fault diagnosis method and system," CN202211368449.4, 2022.

- Z. Yang, F. Liu, and **B. Li**, "A faster-than-real-time observation and analysis method of voltage quality based on fpga," 2022109279359, 2022.

- **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.

- 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.

SKILLS

- **Coding**: Python (proficient), HTML/CSS, JavaScript, C++ — **Frameworks**: PyTorch, Pandas, Spark
- **Languages**: English, Mandarin — **Tools**: Git, Docker, Linux/Unix

ACTIVITIES

- **Academic Service (Reviewer)**: The Web Conference 2024; ACM Transactions on Software Engineering and Methodology (TOSEM); The International Conference on Computer Supported Cooperative Work in Design (CSCWD) 2024.

- **Teaching Assistant**: (CSCI3100) Software Engineering, (CSCI2720) Building Web Applications — 2022-2024
- **Rural Women in STEM Initiative**: Founder — 2022-Present
- **Event Hosting**: PKU Campus Top Ten Signers; PKU International Cultural Exchange Festival — 2018-2019