

Rui Liu

Department of Computer Science, University of Chicago
e: rui Liu@cs.uchicago.edu || w: csruiliu.github.io

Google Scholar, GitHub, LinkedIn

RESEARCH INTERESTS

My current research centers on building systems for machine learning and data science. My previous research experience included building systems for blockchain, data cleaning, mobile computing, and ubiquitous computing.

EDUCATION

Ph.D., Computer Science

University of Chicago Sept, 2017 - Present

- Advised by Prof. Michael J. Franklin and Prof. Aaron J. Elmore, with full scholarship
- Research Topic: Building Multitenant Deep Learning Systems

M.Phil., Computer Science

The Hong Kong Polytechnic University Oct, 2012 - Dec, 2015

- Advised by Prof. Jiannong Cao, with full scholarship
- Thesis: Mitigating Privacy Risks of Smartphones in Mobile Computing

B.Eng., Software Engineering

Northeastern University, China Sept, 2008 - Jun, 2012

- Advised by Prof. Ruiyun Yu, graduate with first honor
- Thesis: Reputation and Incentive Mechanism in Participatory Sensing

EXPERIENCE

Research Assistant

National University of Singapore
Database System Research Group Aug, 2016 - July 2017

- Working with Prof. Beng Chin Ooi
- I am involved in building (a) an evaluation framework for evaluating the performance of private blockchains, (b) a distributed shared memory framework based on RDMA, (c) a data integration and cleaning system for tabular data based on machine learning and crowdsourcing methods.

Research Assistant

The Chinese University of Hong Kong
Laboratory for Applied Security Research Mar, 2016 - Aug, 2016

- Working with Prof. Kehuan Zhang
- I contributed to conducting a comprehensive technical survey on Android apps to investigate the usage of protections, including obfuscation, hidden calls and files.

1. **Multitenancy for Deep Learning in Shared and Resource-Constrained Environments [Ongoing]**

Role: Leader. **Affiliation:** University of Chicago.

- Building a system to support efficient multi-tenant training for deep learning.
- We design a flexible scheduling framework to orchestrate ML training jobs on all available computation resources. The proposed framework embraces the customized reward functions for diverse objectives and generates schedules using reinforcement learning.
- We built a prototype and deployed it to various scenarios like hyper-parameter searching, most-benefit (everyone wants best performance), and sophisticated scenarios (with consideration of various budgets and priority ranking). The results show that the framework can support diverse scenarios and outperforms the existing scheduling mechanisms.

2. **Optimizing Neural Network Training for Hyper-Parameter Tuning**

Role: Leader. **Affiliation:** University of Chicago.

- Building a system to support efficient training for hyper-parameter tuning in deep learning.
- We designed a prototype for a “pack” primitive that allows fine-grained resource sharing in deep learning.
- We built the prototype and evaluated its performance across different models and training scenarios. The results show that there exists a trade-off between packing and unpacking when training multiple neural network models on limited resources.
- We applied the primitive to Hyperband, a widely-used hyper-parameters tuning algorithm. The pack-based Hyperband can be accelerated up to $2.7\times$, with this improvement growing as memory size increases and subsequently the density of models packed.

3. **Draining the Data Swamp**

Role: Core Member. **Affiliation:** University of Chicago.

- Building tools to aid in the taming of data lakes.
- I contributed to implementing a schema completion tool based on Google’s WebTables to evaluate the potential for pure computational solutions to data discovery challenges and investigate our ability to automatically associate schemas.

4. **BlockBench: A Framework for Analyzing Private Blockchains**

Role: Core Member. **Affiliation:** National University of Singapore.

- Building BlockBench, the first evaluation framework for evaluating the performance of private blockchains
- We conducted a comprehensive evaluation of two popular private blockchains: Ethereum and Hyperledger.
- I contributed to building the Hyperledger Fabric (v0.6.0 and v1.0.0) in a private network, implementing APIs, and evaluating the performance.
- I contributed to conducting a comprehensive technical survey on private blockchains using BlockBench.

5. **GAM: Globally Addressable Memory**

Role: Core Member. **Affiliation:** National University of Singapore.

- Building a distributed shared memory framework based on RDMA, which allows users to view the combined memory of a cluster of servers as a single GAM (Globally Addressable Memory)
- I conducted an evaluation for GAM by implementing a key-value store on top of GAM and generating a YCSB-like workload to test the performance.
- I optimized GAM by creating a RDMA-based agent for each node in the cluster. The agent can schedule write and read operations to reduce the overall I/O contention in the cluster.

6. **DICE: A System for Anytime Big Tabular Data Integration and Cleansing**

Role: Core Member. **Affiliation:** National University of Singapore.

- Building a data integration and cleaning system for tabular data based on machine learning and crowdsourcing
- I contributed to developing the DICE-Core, which is used to generate optimal execution plans for different workflows such as schema matching, entity resolution, and data cleaning.

SELECTED PUBLICATIONS

Conference

1. **Rui Liu**, Sanjay Krishnan, Aaron J. Elmore, Michael J. Franklin, “Understanding and Optimizing Packed Neural Network Training for Hyper-Parameter Tuning”, Workshop on Data Management for End-to-End Machine Learning (**DEEM@SIGMOD**) [Long Paper], Virtual, 2021.
2. Heng Zhang, Lingda Li, Donglin Zhuang, **Rui Liu**, Shuang Song, Dingwen Tao, Yanjun Wu, Shuaiwen Leon Song, “An Efficient Uncertain Graph Processing Framework for Heterogeneous Architectures”, ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (**PPoPP**), Virtual, 2021.
3. Will Brackenbury, **Rui Liu**, Mainack Mondal, Aaron J. Elmore, Blase Ur, Kyle Chard, Michael J. Franklin, “Draining the Data Swamp: A Similarity-based Approach”, Workshop on Human-In-the-Loop Data Analytics (**HILDA@SIGMOD**), Houston, USA. 2018.
4. Anh Dinh, Ji Wang, Gang Chen, **Rui Liu**, Beng Chin Ooi, Kian-Lee Tan, “BLOCKBENCH: A Framework for Analyzing Private Blockchains”, ACM International Conference on Management of Data (**SIGMOD**), Chicago, USA. 2017.
5. Zhe Zhou, Wenrui Diao, Xiangyu Liu, Zhou Li, Kehuan Zhang, **Rui Liu**, “Vulnerable GPU Memory Management: Towards Recovering Raw Data from GPU”, Privacy Enhancing Technologies Symposium (**PETS**), Minneapolis, USA. 2017.
6. **Rui Liu**, Jiannong Cao, Sebastian VanSyckel, Wenyu Gao, “PriMe: Human-centric Privacy Measurement based on User Preferences towards Data Sharing in Mobile Participatory Sensing Systems”, IEEE International Conference on Pervasive Computing and Communications (**PerCom**), Sydney, Australia. 2016.

Journal

1. Sanjay Krishnan, Aaron J. Elmore, Michael J. Franklin, John Paparrizos, Zechao Shang, Adam Dziedzic, **Rui Liu**, “Artificial Intelligence in Resource-Constrained and Shared Environments”, ACM SIGOPS Operating Systems Review (**OSR**), vol. 53, no. 1, pp. 1-6, 2019.
2. Ruiyun Yu, Jiannong Cao, **Rui Liu**, Wenyu Gao, Xingwei Wang, Junbin Liang, “Participant Incentive Mechanism towards Quality-Oriented Sensing: Understanding and Application”, ACM Transactions on Sensor Networks (**TOSN**), vol. 15, no. 2, pp. 21:1-21:25, 2019.
3. **Rui Liu**, Junbin Liang, Jiannong Cao, Kehuan Zhang, Wenyu Gao, Lei Yang, Ruiyun Yu, “Understanding Mobile Users’ Privacy Expectations: A Recommendation-based Method through Crowdsourcing”, IEEE Transactions on Services Computing (**TSC**), vol. 12, no. 2, pp. 304-318, 2019.
4. Anh Dinh, **Rui Liu**, Meihui Zhang, Gang Chen, Beng Chin Ooi, Ji Wang, “Untangling Blockchain: A Data Processing View of Blockchain Systems”, IEEE Transactions on Knowledge and Data Engineering (**TKDE**), vol. 30, no. 7, pp. 1366–1385, 2018.
5. **Rui Liu**, Junbin Liang, Jiannong Cao, Weiping Zhu, Sebastian Vansyckel, Christian Becker, “UIO-based Testbed Augmentation for Simulating Cyber-Physical Systems”, **IEEE Intelligent Systems**, vol. 33, no. 6, pp. 69–86, 2018.
6. **Rui Liu**, Jiannong Cao, Kehuan Zhang, Wenyu Gao, Lei Yang, Junbin Liang, “When Privacy Meets Usability: Unobtrusive Privacy Permission Recommendation System for Mobile Apps based on Crowdsourcing”, IEEE Transactions on Services Computing (**TSC**), vol. 11, no. 5, pp. 864–878, 2018.
7. Wenrui Diao, **Rui Liu**, Xiangyu Liu, Zhe Zhou, Zhou Li, Kehuan Zhang, “Accessing mobile user’s privacy based on IME personalization: Understanding and practical attacks”, Journal of Computer Security (**JCS**), vol. 26, no. 3, pp. 283–309, 2018.
8. Junbin Liang, Jiannong Cao, **Rui Liu**, Tao Li, “Distributed Intelligent MEMS: A Survey and a Real-time Programming Framework”, ACM Computing Surveys (**CSUR**), vol. 49, no. 1, pp. 20:1–20:29, 2016.

RECOGNITION

Honors

- 2017: CERES Graduate Scholar Fellowship, granted by UChicago
- 2012: Research Studentship, granted by HK PolyU
- 2012: Outstanding Graduate of Liaoning Province in China (top 1% undergraduate)
- 2011: Google Excellence Scholarship, granted by Google
- 2011: First Class Scholarship, granted by Northeastern University (China)

Awards

- 2020: Technology Award, Department of Computer Science at UChicago
- 2015: Second Runner-up, National Java Programming Contest (China)
- 2013: Honorable Mention, Research Competition of Department of Computing at HK PolyU
- 2010: First Prize, National Undergraduate Mathematical Contest in Modeling (China)
- 2010: Second Runner-up, IBM National Mainframe Technology Application Contest (China)
- 2009: Best Mainframe Application Award, Citi National Finance and Information Technology Contest (China)

Certificates

- 2021: Accelerating CUDA C++ Applications with Multiple GPUs, granted by NVIDIA
- 2021: Fundamentals of Accelerated Computing with CUDA C/C++, granted by NVIDIA

TEACHING

Courses TAed

- 2019, 2020, 2021: CMSC23500/33550, Introduction to Databases, UChicago
- 2018: CMSC11800, Introduction to Data Science, UChicago
- 2014: COMP437, Mobile Computing, HK PolyU
- 2013: COMP320, Introduction to Internet Computing, HK PolyU
- 2013: ENG2003, Information Technology, HK PolyU

REFEREES

Michael J. Franklin

mjfranklin@uchicago.edu

Professor, Liew Family Chairman of Computer Science, University of Chicago

Aaron J. Elmore

aelmore@cs.uchicago.edu

Assistant Professor, University of Chicago

Sanjay Krishnan

skr@uchicago.edu

Assistant Professor, University of Chicago

Beng Chin Ooi

ooibc@comp.nus.edu.sg

Distinguished Professor, Director of Smart Systems Institute, National University of Singapore