# Longbin **Lai**

(+86) 180 6795 0576 | longbin.lai@gmail.com | lai.me | longbinlai | longbin-lai

## Summary

Dr. Longbin Lai is a distinguished expert in distributed systems and graph processing at Alibaba. He has played a crucial role in developing industry-standard and open-source systems such as GraphScope, GAIA, GLogS, and GOpt[1], which enhance interactive graph queries at Alibaba scale. These systems have contributed to GraphScope's record-breaking performance in the latest LDBC SNB benchmark[2] and are widely used in applications like money laundering detection and fraud analysis on Alibaba's platforms. Recently, at Alibaba Tongyi Lab, Dr. Lai is working on large-scale inference engines to improve the performance and reduce the cost of applying large language models.

Dr. Lai earned his Bachelor's and Master's degrees from Shanghai Jiao Tong University (SJTU) in 2010 and 2013, respectively. He completed his Ph.D. at the University of New South Wales (UNSW), Sydney, under the supervision of Prof. Xuemin Lin and Prof. Lu Qin in 2017 and continued his research there as a Postdoctoral Researcher. In 2019, he joined Alibaba Data Academy, where he focuses on developing distributed systems for Alibaba's e-commerce platforms.

## Skills

| | |
|---|---|
| **Programming** | Rust, Python, Java, C/C++, Cypher, Gremlin, SQL, Scala |
| **Big Data** | Hadoop, Spark, Timely dataflow system, Flink, AWS Infrastructure |
| **Big Graph** | GraphScope, Giraph/Pregel, GraphX, Gelly, Neo4J, GraphLab, TigerGraph |
| **Machine Learning** | Tensorflow, PyTorch |
| **Large Language Model** | DeepSpeed, vLLM, LangChain |

## Experience

### Tongyi AI, Alibaba Group — *Hangzhou, China*

Staff Engineer — *Oct. 2023 - Present*

- Product development on LLM applications such as RAG and multi-agent engine,
- Work on the LLM inference systems, covering the topics of memory (KV-cache) management and application-driven scheduling strategies.

### Alibaba Group — *Hangzhou, China*

Staff Engineer — *Dec. 2019 - Oct. 2023*

- The architecturer and product manager of Graph Query Engine, a core component of Alibaba's graph platform: GraphScope (https://github.com/alibaba/graphscope).
- Work on large-scale and intelligent graph query processing system, especially distributed query execution, query compilation and optimization.

### School of Computer Science and Engineering, UNSW — *Sydney, Australia*

Research Assistant — *May. 2017 - May. 2019*

- Design and implement big graph processing primitives and languages.
- Lead a team to develop graph pattern matching system.

### Google Inc. — *Mountain View, CA, USA*

Tech Intern — *Jan. 2017 - Apr. 2017*

- Designed and implemented an emulator that simulates the Google backbone network and the routing strategies for testing, debugging and routing validation.

---

[1] Available at: `https://github.com/alibaba/GraphScope/`

[2] `https://ldbcouncil.org/benchmarks/snb/LDBC_SNB_I_20240514_SF100-300-1000_graphscope.pdf`

### School of Computer Science and Engineering, UNSW

*Sydney, Australia*

PhD candidate, independent research project — *Jul. 2013 - May. 2017*

- (**TwinTwigJoin**) Increased the performance of subgraph enumeration by up to an order of magnitude compared to the state-of-the-art by applying a decomposition-and-join framework in MapReduce.
- (**SEED**) Further improved the TwinTwigJoin by more than one order of magnitude by using a more advanced graph data storage mechanism (extending the traditional adjacency list) and an optimal join structure.

### Department of Advertising and Searching, Alibaba Cloud Computing Corporation

*HangZhou, China*

Research intern, team project — *Jan. 2012 - Sep. 2012*

- Designed and implemented a web recommendation system based on Alibaba cloud computing system (MapReduce-like system), which serves over 1000 top websites in China.
- Improved the throughput of the recommendation system to over 2 billion records per hour via a well-designed MapReduce data flow.
- Implemented a prototype of web classification algorithm that is twice faster than existing algorithm by solely using the url of the web page.

### IBM Share-With-University Project

*Shanghai, China*

Research assistant, project leader — *Oct. 2009 - Oct. 2010*

- Saved the storage overhead of Hadoop File System (HDFS) by up to 30% without compromising the storage reliability by replacing the full replication mechanism with erasure coding.
- Improved the performance of Hadoop streaming utility (allowing coding with languages other than Java) by over 60% by replacing the synchronized inter-process communication module in Linux with desynchronized single-read-single-write queue.

## Education

### The University of New South Wales, Australia (UNSW)

*Sydney, Australia*

PhD. in Computer Science — *Jul. 2013 - Jul. 2017*

- All courses Highly Distinguished

### Shanghai Jiao Tong University (SJTU)

*Shanghai, China*

M.S. in Computer Technology — *Sep. 2010 - Mar. 2013*

- GPA 3.8 / 4.0, China's National scholarship, Top 2%

### Shanghai Jiao Tong University (SJTU)

*Shanghai, China*

B.S. in Information Security — *Sep. 2006 - Jun. 2010*

- GPA 3.6 / 4.0, twice B-class SJTU academic scholarship, Top 15%

## Selected Publications

### Graphy'our Data: Towards End-to-End Modeling, Exploring and Generating Report from Raw Data

*Sigmod Demo 2025, CCF A*

Longbin Lai, Changwei Luo, Yunkai Lou, Mingchen Ju, Zhengyi Yang — *To appear*

### A Modular Graph-Native Query Optimization Framework

*Sigmod 2025, CCF A*

Bingqing Lyu, Xiaoli Zhou, Longbin Lai, Yufan Yang, Yunkai Lou, Wenyuan Yu, Ying Zhang, Jingren Zhou — *To appear*

**Revisiting Graph Analytics Benchmark**

*Sigmod 2025, CCF A*

Lingkai Meng, Yu Shao, Long Yuan, Longbin Lai, Peng Cheng, Xue Li, Wenyuan Yu, Wenjie Zhang, Xuemin Lin, Jingren Zhou

*To appear*

**Towards a Converged Relational-Graph Optimization Framework**

*Sigmod 2025, CCF A*

Yunkai Lou, Longbin Lai, Bingqing Lyu, Yufan Yang, Xiaoli Zhou, Wenyuan Yu, Jingren Zhou

*Berlin, Germany*

**Most Probable Maximum Weighted Butterfly Search**

*ICDE 2025, CCF A*

Yu Shao, Peng Cheng, Longbin Lai, Long Yuan, Wangzi Ni, Xuemin Lin

*To appear*

**A Survey of Distributed Graph Algorithms on Massive Graphs**

*ACM Computing Surveys, CCF A*

Lingkai Meng, Yu Shao, Long Yuan, Longbin Lai, Peng Cheng, Xue Li, Wenyuan Yu, Wenjie Zhang, Xuemin Lin, Jingren Zhou

*Volume 57 Number 2*

**Parallelization of Butterfly Counting on Hierarchical Memory**

*VLDB Journal, CCF A*

Zhibin Wang, Longbin Lai, Yixue Liu, Bing Shui, Chen Tian, Sheng Zhong

*Volume 33, pages 1453–1484*

**GraphScope Flex: LEGO-like Graph Computing Stack**

*SIGMOD 2024, CCF A*

Tao He, Shuxian Hu, Longbin Lai, et al.

*Santiago, Chile*

**GLogS: Interactive Graph Pattern Matching Query At Large Scale**

*ATC 2023, CCF A*

Longbin Lai, Yufan Yang, Zhibin Wang, Yuxuan Liu, Haotian Ma, Sijie Shen, Bingqing Lyu, Xiaoli Zhou, Wenyuan Yu, Zhengping Qian, Chen Tian, Shen Zhong, Yeh-Ching Chung, Jingren Zhou

*Boston, MA, USA*

**Bridging the Gap between Relational OLTP and Graph-based OLAP**

*ATC 2023, CCF A*

Sijie Shen, Zihang Yao, Lin Shi, Lei Wang, Longbin Lai, Qian Tao, Li Su, Rong Chen, Wenyuan Yu, Haibo Chen, Bingyu Zang, Jingren Zhou

*Boston, MA, USA*

**FLASH: A Framework for Programming Distributed Graph Processing Algorithms**

*ICDE 2023, CCF A*

Xue Li, Ke Meng, Lu Qin, Longbin Lai, Wenyuan Yu, Zhengping Qian, Xuemin Lin, Jingren Zhou

*Anaheim, CA, USA*

**I/O-Efficient Butterfly Counting at Scale**

*Sigmod 2023, CCF A*

Zhibin Wang, Longbin Lai, Yixue Liu, Bing Shui, Chen Tian, Sheng Zhong

*Seattle, WA, USA*

**GraphScope: A Unified Engine For Big Graph Processing**

*VLDB 2021, CCF A*

Wenfei Fan, Tao He, Longbin Lai, Xue Li, Yong Li, Zhao Li, Zhengping Qian, Chao Tian, Lei Wang, Jingbo Xu, Youyang Yao, Qiang Yin, Wenyuan Yu, Jingren Zhou, Diwen Zhu, Rong Zhu

*PVLDB Volume 14, Number 3*

**HUGE: An Efficient and Scalable Subgraph Enumeration System**

*Sigmod 2021, CCF A*

Zhengyi Yang, Longbin Lai, Xuemin Lin, Kongzhang Hao, Wenjie Zhang

**GAIA: A System for Interactive Analysis on Distributed Graphs Using a High-Level Language** — *NSDI 2021, CCF A*

Zhengping Qian, Chenqiang Min, Longbin Lai, Yong Fang, Gaofeng Li, Youyang Yao, Bingqing Lyu, Zhimin Chen, Jingren Zhou

**Efficient structural node similarity computation on billion-scale graphs** — *VLDB Journal, CCF A*

Xiaoshuang Chen, Longbin Lai, Lu Qin, Xuemin Lin

**A Framework to Quantify Approximate Simulation on Graph Data** — *ICDE 2021, CCF A*

Xiaoshuang Chen, Longbin Lai, Lu Qin, Xuemin Lin, Boge Liu

**Distributed Subgraph Matching on Timely Dataflow** — *VLDB 2019, CCF A*

Longbin Lai, Zhu Qing, Zhengyi Yang, Xin Jin, Zhengmin Lai, Ran Wang, Kongzhang Hao, Xuemin Lin, Lu Qin, Wenjie Zhang, Ying Zhang, Zhengping Qian, Jingren Zhou — *PVLDB Volume 12, Number 10*

**Scalable Distributed Subgraph Enumeration** — *VLDB 2017, CCF A*

Longbin Lai, Lu Qin, Xuemin Lin, Ying Zhang, Lijun Chang, Shiyu Yang — *PVLDB Volume 10, Number 3*

**Scalable Subgraph Enumeration in MapReduce, A Cost-oriented Approach** — *VLDB Journal, CCF A*

Longbin Lai, Lu Qin, Xuemmin Lin, Lijun Chang — *Volume 26, Number 3*

**Scalable Subgraph Enumeration in MapReduce** — *VLDB 2015, CCF A*

Longbin Lai, Lu Qin, Xuemmin Lin, Lijun Chang — *PVLDB Volume 8, Number 10*

## Honors & Awards

| | | |
|---|---|---|
| 2021 | **C-class Talent**, HangZhou 521 Project of Talent Introduction | *Hangzhou, China* |
| 2012 | **Top 1%**, China's National Scholarship | *SJTU, China* |
| 2011 | **Top 4%**, Tencent Academic Scholarship | *SJTU, China* |
| 2010 | **Top 10%**, Outstanding Graduate of Shanghai Jiao Tong University | *SJTU, China* |
| 2009 | **Top 6%**, Sony Academic Scholarship | *SJTU, China* |
| 07, 08 | **Top 15%**, B-Class SJTU Academic Scholarship | *SJTU, China* |