Bo Zhao | Assistant Professor

Department of Computer Science, **Aalto University**
Konemiehentie 2, 02150 Espoo, Finland
zbjob.github.io • Email: bo.zhao@aalto.fi

Research Interest

I conduct research on efficient ***machine learning (ML) systems*** that translate ***data*** into ***value*** for decision making. The scope of my research spans across multiple subfields, from **scalable reinforcement learning systems** to **distributed state and data management systems**, as well as **code optimization** techniques. That is to answer the question ***"how to co-design multiple layers of the software stack to improve the scalability, performance, and energy efficiency of ML systems"***. My long-term goal is to understand the fundamental connections between data/knowledge management and modern ML systems to make decision-making more transparent, robust and efficient.

Education

PhD in Computer Science — **Berlin, Germany**
Humboldt-Universität zu Berlin — *02/2016–05/2022*
Thesis: State Management for Efficient Event Pattern Detection
Supervisor: Prof. Dr. Matthias Weidlich
Honors: *magna cum laude*

M.Sc. in Computer Science — **Xi'an, China**
Xi'an Jiaotong University — *09/2012–07/2015*
Thesis: Dependence-Based Coarse-Grained Automatic Parallelisation
Ranking: Top 1% of the university
Honors: *summa cum laude*

B.Sc in Computer Science — **Wuhan, China**
Wuhan Institute of Technology — *09/2008–07/2012*
Thesis: Energy-Aware Routing Optimizations in Wireless Sensor Networks
Ranking: Top 1% of the university
Honors: *summa cum laude*

Visiting Study

University of Queensland — **Brisbane, Australia**
Visiting PhD Student in Computer Science, hosted by Prof. Xiaofang Zhou — *05/2017–06/2017*
Topic: Efficient Data Stream Processing

RWTH-AACHEN University — **Aachen, Germany**
Visiting M.Sc. Student in Computer Science, hosted by Prof. Felix Wolf — *09/2013–02/2015*
Topic: High Performance Computing

Work Experience

09/2023-present: Assistant Professor, **Aalto University, Espoo, Finland**

01/2023-08/2023: Assistant Professor, **Queen Mary University of London, London, UK**
Honorary Research Fellow, **Imperial College London, London, UK**

07/2021–01/2023: Postdoctoral Researcher, **Imperial College London, London, UK**

02/2016–06/2021: Doctoral Researcher in **Humboldt-Universität zu Berlin, Berlin, Germany**

06/2019–09/2019: Software Development Engineer Intern at Amazon, **AWS Redshift, Berlin, Germany**

11/2015–01/2016: Research Assistant in **Technische Universität Darmstadt, Darmstadt, Germany**

10/2013–02/2015: Student Research Assistant in **RWTH-AACHEN University, Aachen, Germany**

Grants / Projects

AthenaRL: Scalable and Flexible Distributed Reinforcement Learning Systems (link)
Funding agency: Research Council of Finland — Academy Project Funding
Duration: 2024-2028
Role: Principal investigator (PI)
Amount: 780,114 EUR

AthenaQEC: Real-Time Decoding of over 1,000 Logical Qubits
Funding agency: Research Council of Finland–Finnish Quantum Flagship Exploratory Projects: Quantum Future (P4)

Duration: 2025-2027
Role: Principal investigator (PI)
Amount: 75,000 EUR

LARA: Large Language Models for Quantum Machine Learning Algorithms
Funding agency: Business Finland — Quantum Computing Research Call
Duration: 2024-2026
Role: Principal investigator (PI)
Amount: 349,915 EUR (out of 700,000 EUR)

FlexMoE: Flexible Efficient Mixture-of-Experts Systems
Funding agency: The Finnish Doctoral Program Network in Artificial Intelligence
Duration: 2024-2027
Role: Principal investigator (PI)
Amount: 112,011 EUR

LashQ: Large Language Model Augmented Scalable Hybrid Quantum-Classical Computing Framework
Funding agency: Research Council of Finland–the Finnish Quantum Flagship's Quantum Doctoral Pilot Programme
Duration: 2024-2027
Role: Principal investigator (PI)
Amount: 112,011 EUR

CloudButton: a Serverless Data Analytics Platform
Funding agency: EU Horizon 2020 Framework Programme
Duration: 2019-2022
Role: Participant (Postdoctoral researcher)
Amount: 4.2 million EUR

Process-Awareness of Event-Driven Systems: Model, Analysis and Optimisation
Funding agency: German Research Foundation (Deutsche Forschungsgemeinschaft, DFG)
Duration: 2014-2021
Role: Participant (Doctoral researcher)
Amount: 1 million EUR

Honors & Awards

2024-2025: VLDB Distinguished Reviewer Award

2023-2024: The 4th Best Course for the Department of Computer Science, Aalto University

2019-2020: Travel Grant of the Silk Road International Symposium for Distinguished Young Scholars

2017–2018: IEEE ICDE Student Travel Grant

2015–2016: EDBT Summer School Travel Grant

2014–2015: Outstanding Graduate, ACM SIGPLAN Travel Grant, ACM SIGMICRO Travel Grant

2012–2013: China National Scholarship(top 0.2%), Creative-Master Scholarship, Excellent Master Student

2011–2012: Excellent Graduation Thesis, Outstanding Graduate

2010–2011: China National Scholarship(top 0.2%), Top Grade Scholarship, Pacemaker to Merit Student, Advanced Individual in Social Practice

2009–2010: China National Scholarship(top 0.2%), Top Grade Scholarship, Pacemaker to Merit Student

2008–2009: Top Grade Scholarship, Pacemaker to Merit Student, Outstanding League Member

Publications

Google Scholar Profile DBLP Profile

- Alessandro Fogli, **Bo Zhao**, Peter Pietzuch, Jana Giceva: **CHARM: Chiplet Heterogeneity-Aware Runtime Mapping System**, European Conference on Computer Systems (**EuroSys'26**), Edinburgh, UK, 2026 (to appear)

- Zheyue Tan, Mustapha Abdullahi, Tuo Shi, Huining Yuan, Zelai Xu, Chao Yu, Boxun Li, **Bo Zhao**: **EARL: Efficient Agentic Reinforcement Learning Systems for Large Language Models**, Workshop on Systems for Agentic AI (SAA), in conjunction with the ACM Symposium on Operating Systems Principles (**SOSP'25**), Seoul, Korea, 2025 (to appear)

- Cong Yu, Tuo Shi, Matthias Weidlich, **Bo Zhao**: **SHARP: Shared State Reduction for Efficient Matching of Sequential Patterns**, *preprint, 2025*

- Boxun Li, Yadong Li, Zhiyuan Li, Congyi Liu, Weilin Liu, Guowei Niu, Zheyue Tan, Haiyang Xu, Zhuyu Yao, Tao Yuan, Dong Zhou, Yueqing Zhuang, **Bo Zhao**, Guohao Dai, Yu Wang: **Megrez2 Technical Report**, *2025*

- Linus Jern, Valter Uotila, Cong Yu, **Bo Zhao**: **Agent-Q: Fine-Tuning Large Language Models for Quantum Circuit Generation and Optimization**, *In the Proc. of IEEE International Conference on Quantum Computing and Engineering (**QCE'25**), Albuquerque, NM, USA, 2025*

- Valter Uotila, Julia Ripatti, **Bo Zhao**: **Higher-Order Portfolio Optimization with Quantum Approximate Optimization Algorithm**, *In the Proc. of IEEE International Conference on Quantum Computing and Engineering (**QCE'25**), Albuquerque, NM, USA, 2025*

- Lei You, Lele Cao, Mattias Nilsson, **Bo Zhao**, Lei Lei: **Distributional Counterfactual Explanation With Optimal Transport**, *In the Proc. of International Conference on Artificial Intelligence and Statistics (**AISTATS'25**) (Oral, top 2%), Thailand, 2025*

- Alessandro Fogli, **Bo Zhao**, Peter Pietzuch, Jana Giceva: **ARCAS: Adaptive Runtime System for Chiplet-Aware Scheduling**, *preprint, 2025*

- Song Liu, Jie Ma, Zhenyuan Zhang, Xinhe Wan, **Bo Zhao**, Weiguo Wu: **Scalpel: High Performance Contention-Aware Task Co-Scheduling for Shared Cache Hierarchy**, *In IEEE Transactions on Computers, Volume. 74, 2025*

- Marcel Wagenländer, Guo Li, **Bo Zhao**, Luo Mai, Peter Pietzuch: **TENPLEX: Changing Resources of Deep Learning Jobs using Parallelizable Tensor Collections**, *In the Proc. of Symposium on Operating Systems Principles (**SOSP'24**), Austin, TX, USA, 2024*

- Alessandro Fogli, **Bo Zhao**, Peter Pietzuch, Maximilian Bandle, Jana Giceva: **OLAP on Modern Chiplet-Based Processors**, *In the Proc. of International Conference on Very Large Data Bases (**VLDB'24**), Guangzhou, China, 2024*

- Huanzhou Zhu*, **Bo Zhao***, Gang Chen, Weifeng Chen, Yijie Chen, Liang Shi, Yaodong Yang, Peter Pietzuch, Lei Chen (***equal contribution**): **MSRL: Distributed reinforcement learning with dataflow fragments**, *In Proc. of the USENIX Annual Technical Conference (**USENIX ATC'23**), Boston, MA, USA, July, 2023*

- Song Liu, Xinhe Wan, Zengyuan Zhang, **Bo Zhao**, Weiguo Wu: **TurboStencil: You Only Compute Once for Stencil Computation**, ***Future Generation Computer Systems** (IF=7.307), 2023*

- Gururaghav Raman, **Bo Zhao**, Jimmy Chih-Hsien Peng, Matthias Weidlich: **Adaptive incentive-based demand response with distributed non-compliance assessment**, ***Applied Energy*** (IF=11.446), Volume 326, November, 2022

- **Bo Zhao** : **State Management for Efficient Event Pattern Detection**, *Dissertation*, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, May 2022

- **Bo Zhao**, Han van der Aa, Nguyen Thanh Tam, Nguyen Quoc Viet Hung, Matthias Weidlich: **EIRES: Efficient Integration of Remote Data in Event Stream Processing**, *In Proc. of the 47th ACM SIGMOD International Conference on Management of Data (**SIGMOD'21**), Xi'an, China, ACM, June 2021*

- **Bo Zhao**, Nguyen Quoc Viet Hung, Matthias Weidlich: **Load Shedding for Complex Event Processing: Input-based and State-based Techniques**, *In Proc. of the 36th IEEE International Conference on Data Engineering (**ICDE'20**), Dallas, TX, USA, IEEE, April 2020*

- Gururaghav Raman, Jimmy Chih-Hsien Peng, **Bo Zhao**, Matthias Weidlich: **Dynamic Decision Making for Demand Response through Adaptive Event Stream Monitoring**, *In Proc. of the IEEE Power & Energy Society General Meeting (**PESGM'19**), Atlanta, GA, USA. IEEE, August 2019.*

- **Bo Zhao**: **Complex Event Processing under Constrained Resources by State-based Load Shedding**, *In Proc. of the 34th IEEE International Conference on Data Engineering (**ICDE'18**), Paris, France, IEEE, April 2018*

- **Bo Zhao**, Zhen Li, Ali Jannesari, Felix Wolf, Weiguo Wu: **Dependence-Based Code Transformation for Coarse-Grained Parallelism**, *In Proc. of the International Workshop on Code Optimisation for Multi and Many Cores (**COSMIC'15**) held in conjunction with **CGO'15**, San Francisco Bay Area, CA, USA, ACM, February 2015*

- **Bo Zhao**, Ali Jannesari: **Dependence-Based Parallel Code Generation Using Intel CnC**, *In Proc. of the 24th International Conference on Parallel Architectures and Compilation Techniques (**PACT'15**),San Francisco Bay Area, CA, USA, October 2015* (ACM SRC poster)

- Zhen Li, **Bo Zhao**, Ali Jannesari, Felix Wolf: **Beyond Data Parallelism: Identifying Parallel Tasks in Sequential Programs**, *In Proc. of the 15th International Conference on Algorithms and Architectures for Parallel Processing (**ICA3PP'15**), Springer, November 2015*

Research Talks

June 2025: Invited talk at Technical University of Munich, Munich, Germany;
April 2025: Invited talk at ByteDance Ltd., Virtual Event, San Jose, CA, USA;
March 2025: Invited talk at the Berlin Institute for the Foundations of Learning and Data (BIFOLD), Berlin, Germany;
January 2025: Invited talk at Shanghai Jiaotong University, Shanghai, China;
January 2025: Invited talk at Beijing Institute of Technology, Beijing, China;
May 2024: Invited guest lecture at TU Wien, Vienna, Austria;
January 2024: Invited talk at Peking University, Virtual Event, China;
November 2023: Invited talk at AI Day–Finnish Center for Artificial Intelligence, Espoo, Finland;
June 2023: Invited guest lecture at Humboldt-Universität zu Berlin, Berlin, Germany;
June 2023: Invited talk at the Huawei Cloud InnovWave Overseas Workshop, Munich, Germany;
June 2023: Invited guest lecture at TU Wien, Vienna, Austria;
May 2023: USENIX Annual Technical Conference (*USENIX ATC'23*), Boston, MA, USA;
May 2023: Invited talk at the Global Software Technology Summit, Dresden, Germany;
May 2023: Invited talk at the Max Planck Institute for Software Systems (MPI-SWS), Saarbrücken, Germany;
March 2023: Invited talk at Aalto University, Espoo, Finland;
January 2023: Invited talk at TU Wien, Vienna, Austria;
November 2022: Invited talk at King's College London, London, UK;
December 2021: Invited talk at Xi'an Jiaotong University, Virtual Event, China;
November 2021: Invited talk at Nanjing University, Virtual Event, China;
June 2021: The 47th ACM International Conference on Management of Data (*SIGMOD'21*), Virtual Event, China;
March 2021: Invited talk at EPFL, Lausanne, Switzerland;
December 2020: Invited talk at Hasso Plattner Institute, Potsdam, Germany;
November 2020: Invited talk at ETH Zürich, Zürich, Switzerland;
November 2020: Invited talk at Imperial College London, London, UK;
November 2020: Invited talk at Technical University of Berlin, Berlin Germany;
April 2020: The 36th IEEE International Conference on Data Engineering (*ICDE'20*), Dallas, TX, USA;
April 2019: Invited talk at Xi'an Jiaotong University, Xi'an, China;
April 2018: The 34th IEEE International Conference on Data Engineering (*ICDE'18*), Paris, France;
September 2015: The 7th Annual Concurrent Collections Workshop (*with LCPC'15*), Raleigh, NC, USA;
September 2015: The 44th International Conference on Parallel Processing (*ICPP'15*), Beijing, China;
February 2015: The 2nd International Workshop on Code Optimisation for Multi and Many Cores (*COSMIC'15*), San Francisco Bay Area, CA, USA;
September 2014: The Sixth Annual Concurrent Collections Workshop, Intel Corp in Hillsboro, OR, USA;

Academic Services

Organizing Committees: Local Arrangement Chair of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (***UbiComp***) 2025

Program Committees: European Conference on Computer Systems (***EuroSys***) 2026
International Conference on Management of Data (***SIGMOD***) 2026
International Conference on Very Large Data Bases (***VLDB***) 2025, 2026
The ACM International Conference on Emerging Networking Experiments and Technologies (***CoNEXT***) 2024, 2025
IEEE International Conference on Data Engineering (***ICDE***) 2025
The ACM Conference on Information and Knowledge Management (***CIKM***) 2021, 2022, 2023

Availability & Reproducibility Committees: The ACM International Conference on Management of Data (***SIGMOD***) 2022, 2023

Demo Track Program Committees: IEEE International Conference on Data Engineering (***ICDE***) 2023, 2024, 2025

Reviewers for Journals: IEEE Transactions on Computers 2025
The VLDB Journal (***VLDBJ***) 2025
The Journal of Machine Learning Research (***JMLR***)) 2024
IEEE Transactions on Parallel and Distributed Systems (***TPDS***) 2023
Journal of Systems and Software (***JSS***) 2016

Journal Editor: Proceedings of the ACM on Networking (***PACMNET***) 2024

Teaching Experience

Autumn 2025: CS-E4780 *Scalable Systems and Data Management*, Aalto University

Spring 2025: CS-E4645 *Research Project on Data Intensive Computing*, Aalto University

Autumn 2024: CS-E4780 *Scalable Systems and Data Management*, Aalto University

Autumn 2023: CS-E4190 *Cloud Software and Systems*, Aalto University (co-teaching with Prof. Mario Di Francesco)

Semester B 2023: ECS656U *Distributed Systems*, Queen Mary University of London

Summer semester 2020: Seminar on *Distributed Data Management Systems*, Humboldt-Universität zu Berlin

Summer semester 2020: Oral exam examiner on *Process Mining*, Humboldt-Universität zu Berlin

Winter semester 2019: Oral exam examiner on *Event Process*, Humboldt-Universität zu Berlin

Winter semester 2019: Exercises (Übung) on *Data Stream Processing*, Humboldt-Universität zu Berlin

Summer semester 2018: Oral exam examiner on *Process Mining*, Humboldt-Universität zu Berlin

Summer semester 2018: Seminar on *Event Stream Processing*, Humboldt-Universität zu Berlin

Student Mentoring

- PhD dissertation on "Efficient and Adaptive Machine Learning Data Pipeline" (tentative), Ms. Jiaxin Guo, June. 2024-present, Aalto University, Finland
- PhD dissertation on "Efficient State Management for Retrieval-Augmented Large Language Models" (tentative), Mr. Alireza Samar, April 2024-present, Aalto University, Finland
- PhD dissertation on "Scalable Reinforcement Learning Systems on Supercomputers" (tentative), Mr. Mustapha Abdullahi, February 2024-present, Aalto University, Finland
- PhD dissertation on "Scalable and Flexible Machine Learning Pipelines" (tentative), Mr. Cong Yu, December 2023-present, Aalto University, Finland
- Master thesis project on "Efficient GPU Resource Utilization Monitoring on the LUMI Supercomputer ", Mr. Songlin Jiang, December 2023-present, Aalto University, Finland
- Master thesis project on "Data-Mesh-Enhanced Tekla Structures Environment Management", Mr. Xu Feng, December 2023-present, Aalto University, Finland
- Master thesis project on "Dataflow-Based MLOps for Machine Learning Pipelines", Mr. Vishnu Puramchalil, December 2022-August 2023, Queen Mary University of London, UK
- Master thesis project on "Adaptive Query Analytics over Dynamic Data Streams", Mr. Shaurya Rana, December 2022-August 2023, Queen Mary University of London, UK
- Master thesis project on "Dataflow Optimisation for Scalable Reinforcement Learning Systems", Mr. Mustapha Abdullahi, December 2022-August 2023, Queen Mary University of London, UK
- Master thesis project on "Distributed Data Stream Processing for Business Intelligence", Mr. Chinar Amrutkar, December 2022-August 2023, Queen Mary University of London, UK
- Student intern project on "Implementing MuZero Agents Using Mindspore Computation Graphs", Mr. Liyi Tan, June 2022-September 2022, Imperial College London, UK
- Undergraduate project on "Implementing MuZero Algorithm Using the Mindspore DL Engine", Mr. Bartłomiej Cieślar, January 2022-May 2022, Imperial College London, UK
- Master thesis project on "Mining Constraints to Optimize CEP Load Shedding for Multiple Queries", Mr. Xudong Zhu, 2019-2020, Humboldt-Universität zu Berlin, Germany

References (with website links on names)

Name	Affiliation	Email address
Prof. Peter Pietzuch	Imperial College London, UK	prp@imperial.ac.uk
Prof. Matthias Weidlich	Humboldt-Universität zu Berlin, Germany	matthias.weidlich@hu-berlin.de
Prof. Ivona Brandić	Vienna University of Technology, Austria	ivona.brandic@tuwien.ac.at
Prof. Nguyen Quoc Viet Hung	Griffith University, Australia	henry.nguyen@griffith.edu.au
Prof. Han van der Aa	University of Vienna, Austria	han.van.der.aa@univie.ac.at