

Yanlin Qi

Université Paris Cité, Paris, France

E-mail: yanlinqi7@gmail.com

HomePage: <https://amy-77.github.io>

Tel: +33-744945844

EDUCATION

Université Paris Cité (PhD Candidate)

2024.12-Now

- Supervisor: **Prof. Themis Palpanas** (ACM Fellow) Major: Computer Science
- Research direction: KV Cache Retrieval, Efficient LLM, Vector Similarity Search

Harbin Institute of Technology, Shenzhen (Master)

2021.09-2024.01

- Supervisor: **Prof. Guoting Chen** Major: Mathematics
- Research direction: Pattern mining GPA: 3.42 / 4

Dalian Maritime University (Bachelor)

2017.09-2021.06

- GPA: 92.9/100 Ranking: 1/59 Major: Statistics

PUBLICATION

- **ParisKV: Fast and Drift-Robust KV-Cache Retrieval for Long-Context LLMs.** Yanlin Qi, Xinhang Chen, Huiqiang Jiang, Qitong Wang, Botao Peng, Themis Palpanas. (Under review) 2026.02
- **GoodLeaf: Confidence-Calibrated Learned Filters for Data Series Indexing.** Yanlin Qi, Qitong Wang, Themis Palpanas. (To be submitted) 2026.02
- **LEAD: Iterative Data Selection for Efficient LLM Instruction Tuning.** Xiaotian Lin, Yanlin Qi, Yizhang Zhu, Themis Palpanas, Nan Tang, Yuyu Luo. *VLDB 2026*
- **TokenTune: Dual-Level Utility Estimation for Scalable Data Selection In Instruction Tuning.** Xiaotian Lin, Yanlin Qi, Yuxiang Luo, Themis Palpanas, Chengliang Chai, Yuyu Luo (Under review) 2026.02
- **Why Are Learned Indexes So Effective but Sometimes Ineffective?** Qiyu Liu, Siyuan Han, Yanlin Qi*(co-first), Jingshu Peng, Jin Li, Longlong Lin, Lei Chen. *VLDB 2025*
- **Not Small Enough? SegPO: Learning to Further Compress Product Quantization Codebooks.** Qiyu Liu, Yanlin Qi*(co-first), Siyuan Han, Jingshu Peng, Jin Li, Lei Chen. *VLDB 2025*
- **TokenFlow: Zero-Cost Token Selection for Data-Efficient LLM Fine-Tuning.** Xiaotian Lin, Zhuowen Liang, Yanlin Qi, Yuyu Luo (Under review) 2025.10
- **RecRanker: Instruction Tuning Large Language Model as Ranker for Top-k Recommendation [J].** ACM Transactions on Information Systems. Sichun Luo, Bowei He, Haohan Zhao, Wei Shao, Yanlin Qi, et al. 2024.08
- **Mining periodic trends via closed high utility patterns[J].** Yanlin Qi, Xiaojie Zhang, Wensheng Gan, Guoting Chen. *Expert Systems with Applications, SCI I, 2023, IF: 8.665*
- **F-RFM-Miner: An efficient algorithm for mining valuable fuzzy patterns via the RFM model[J].** Yanlin Qi, Fuyin Lai, Guoting Chen, Wensheng Gan. *Applied Intelligence, 2023, SCI II, IF:5.3*
- **Fuzzy-driven periodic frequent pattern mining [J].** Xiaojie Zhang, Yanlin Qi, Guoting Chen, Wensheng Gan, Philippe Fournier-Viger. *Information Sciences, 2022, SCI I, IF: 8.233*

- [Mining Valuable Fuzzy Patterns via the RFM Model \[C\]](#). Yanlin Qi, Guoting Chen, Wensheng Gan. (ICDMW). 2022

EXPERIENCE

- Hong Kong University of Science and Technology (Visiting Student)** 2023.06-2024.01
- Supervisor: **Lei Chen** (ACM Fellow) **Major: Computer Science**
 - Research direction: AI for DB : learned index, product quantization, lossless compression.
- City University of Hong Kong (Research Assistant)** 2024.01-2024.08
- Supervisor: **Zhichao Lu** **Major: Computer Science**
 - Research direction: Automatic Algorithm Design, Large Language Model + Evolutionary Computation
- Hong Kong University of Science and Technology(Guangzhou) (RA)** 2024.09-2024.12
- Supervisor: **Yuyu Luo** **Major: Computer Science**
 - Research direction: LLM fine-tune for Data Selection

Research Interests

- My previous research direction is Pattern mining, LLM for Automatic Algorithm design, data selection.
- Now I focus on LLM inference optimization, KV cache Retrieval, Token Pruning
- Skills: Strong programming ability: C++ =Java > Python ≥ Matlab > R > SPSS; Strong mathematical ability.
- Languages: Chinese (native), English (fluent) **Github:** <https://github.com/amy-77>

Project Experiences

1. **DataGEM (EOSC)** [Stream Data Anomaly Detection \(Contributor\)](#) 2025.10-Now
Contributed to the open-source DataGEMS ecosystem by **designing and implementing a real-time anomaly detection module** for streaming data, end-to-end from method design to code integration and evaluation.
2. [Multi-Objective Optimization for Intelligent Resource Management in Big Data Processing Systemc](#) 2023.06
Big data processing at the production scale presents a highly complex environment for resource optimization (RO). We improved a **stage-level MOO** problem into a **query-level MOO** problem.
3. [Knowledge Discovery on Multi-temporal Data](#) | *Shenzhen Basic Research Project* 2021.10-2023.01
We aim to discover the relationship between items hidden in the database through association rules. We introduce the fuzzy dataset, introduce the negative-utility dataset, and propose some special **pattern mining algorithms**

HONORS & AWARDS

- **National Scholarship (3)** 2018.06-2021.06
- **First-class scholarship of Harbin Institute of Technology(3)** 2021.09-2022.09
- **Excellent master's thesis of Harbin Institute of Technology (2%)** 2024.01
- **2020 College Students Innovation and Entrepreneurship Training Program (2)** 2019.09-2020.12
- **Excellent student** of Dalian Maritime University (2%) 2021.05
- The third prize in **Dalian Municipal Mathematics Contest** (5%) 2019.05
- The second prize in **Mathematics Competition** of Dalian Maritime University (5%) 2019.05