

Associate Professor (Special Appointment)

+86-13430470235

Large Language Model Research Center

chenlei@suat-sz.edu.cn

Artificial Intelligence Research Institute

<https://jasonclei.github.io>

Shenzhen University of Advanced Technology (SUAT)

Google Scholar: <https://scholar.google.com/citations?hl=zh-CN&user=H25qJSMAAAAJ>

Summary of achievements

- Published over 25 papers in peer-reviewed journals and conferences (Google Scholar statistics: Citations (1,000+), h-index (12)), including *Nature Methods*, *Nature Communications*, *IEEE Transactions on Knowledge and Data Engineering*, *Genome Research*, *NeurIPS*, *AAAI*, *SIGIR*, *EMNLP*, *ACM BCB* etc.
- Defined the current global performance standard of nanopore long-read small variant calling in competition with Google Health. My solutions are published in *Nature Methods*, *Nature Communications*, and *Genome Research*, and are being used routinely in major bioinformatics centers and medical centers internationally, and locally.
- Awards and recognitions:
 - 2026, HKU Foundation Publication Award for Research Postgraduate Students
 - 2022, Shenzhen Universiade International Scholarship
 - 2021, Tencent Rhino-Bird Research Elite Training Program
 - 2022, Breakthrough Progress Award, Tencent
 - 2021, 2022, Dean Scholarship-Excellent & Special Award, SIAT
 - 2019, 2022, Outstanding Graduates & Graduates Thesis
 - 2018, National Scholarship for Encouragement
 - 2016, National Scholarship
- High-impact open-source software:
 - ClairS, ClairS-TO, Clair-Mosaic, Clair3-RNA, Gungnir, etc.

Research interests

1. AI for Science

2. Natural Language Processing

Education

Ph.D. in Computer Science

Sept 2022 – Sept 2026

The University of Hong Kong, Hong Kong

Supervisor: Prof. LUO, Ruibang

M.Eng. in Computer Technology

Sept 2019 – Jun 2022

University of Chinese Academy of Sciences, Beijing

Affiliation: Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

Supervisor: Prof. YANG, Min

B.Eng. in Computer Science and Technology

Sept 2015 – Jun 2019

China University of Mining and Technology, Xuzhou

Experience

Shenzhen University of Advanced Technology (SUAT)

Oct 2026 – Now

Associate Professor (Special Appointment)

Large Language Model Research Center, Artificial Intelligence Research Institute

WeiXin Group (WXG), Tencent

Jun 2021 – Jun 2022

2021 Tencent Rhino-Bird Research Elite Training Program, Research intern

Advisor: Prof. DING, Jingtao

Westlake University

Apr 2021 – Aug 2021

Research intern

Advisor: Prof. YUAN, Fajie

Shenzhen Institute of Advanced Technology, CAS

Sept 2018 – Aug 2019

Research intern

Advisor: Prof. YANG, Min

Awards and honors

1. 2026, HKU Foundation Publication Award for Research Postgraduate Students
2. 2022, Shenzhen Universiade International Scholarship
3. 2021, Tencent Rhino-Bird Research Elite Training Program
4. 2022, Breakthrough Progress Award, Tencent
5. 2021, 2022, Dean Scholarship-Excellent & Special Award, SIAT
6. 2019, 2022, Outstanding Graduates Thesis & Graduates
7. 2018, National Scholarship for Encouragement

8. 2016, National Scholarship

Peer-reviewed publications (as of October 2026)

In my following listing, my publications are grouped into two categories:

- First-authored (^) and Co-first authored (#) publications
- Co-authored publications

Impact factors are as of October 2026. The full list of my publications is arranged and easily accessible from my Google Scholar profile at

<https://scholar.google.com/citations?user=H25qJSMAAAAI>

First-authored (^) and Co-first authored (#) publications:

1. L. Chen[^], Z. Zheng, J. Su, X. Yu, A. Wong, J. Zhang, Y. Lee, R. Luo. ClairS-TO: a deep-learning method for long-read tumor-only somatic small variant calling. *Nature Communications (IF: 18.10)*, 16(1), 9630, 2025.
2. Z. Zheng, L. Chen[#], J. Su, X. Yu, M. He, Y. Lee, T. Lam, R. Luo. ClairS: a deep-learning method for long-read tumor-normal pair somatic small variant calling. *Nature Methods (IF: 28.30)*, 2026.
3. J. Zhang, L. Chen[#], J. Sun, S. Li, Y. Zhou, ..., C. Li, Z. Zheng, R. Luo. Gungnir codec enabling high error-tolerance and low-redundancy DNA storage through substantial computing power. *Nature Communications (IF: 18.10)*, 2026.
4. L. Chen[^], Z. Zheng, M. He, A. Wong, X. Yu, J. Li, J. Zhang, Y. Zhou, R. Luo. Accurate long-read mosaic small variant calling with Clair-Mosaic. *Genome Research (IF: 6.30)*, 2026.
5. L. Chen[^], F. Yuan, J. Yang, X. He, C. Li, M. Yang. User-specific adaptive fine-tuning for cross-domain recommendations. *IEEE Transactions on Knowledge and Data Engineering (IF: 11.60)*, 35(3), 3239-3252, 2021.
6. L. Chen[^], F. Yuan, J. Yang, X. Ao, C. Li, M. Yang. A user-adaptive layer selection framework for very deep sequential recommender models. *AAAI (CCF A)* (Vol. 35, No. 5, pp. 3984-3991), 2021.
7. L. Chen[^], J. Ding, M. Yang, C. Li, C. Song, L. Yi. Item-provider co-learning for sequential recommendation. *SIGIR (CCF A)* (pp. 1817-1822), 2022.
8. L. Chen[^], F. Yuan, J. Yang, C. Li, M. Yang. Scene-adaptive Knowledge Distillation for Sequential Recommendation via Differentiable Architecture Search. *WANT@NeurIPS Workshop*, 2023.

9. **L. Chen[^]**, J. Su, Z. Zheng, T. Lam, R. Luo. Large-scale Dataset and Effective Model for Variant-Disease Associations Extraction. *ACM BCB*, 10.1145/3584371.3612995, 2023.
10. **L. Chen[^]**, J. Su, T. Lam, R. Luo. Exploring Pair-Aware Triangular Attention for Biomedical Relation Extraction. *ACM BCB*, 10.1145/3584371.3612994, 2023.
11. **L. Chen[^]**, R. Xu, M. Yang. Overview of the NLPCC 2020 shared task: multi-aspect-based multi-sentiment analysis (MAMS). *NLPCC (CCF C)* (pp. 579-585). Cham: Springer International Publishing. 2020.

Co-authored publications:

12. Z. Zheng, X. Yu, **L. Chen**, Y. Lee, C. Xin, ..., R. Kesharwani, F. Sedlazeck, R. Luo. Clair3-RNA: A deep learning-based small variant caller for long-read RNA sequencing data. *Nature Communications (IF: 18.10)*, 16(1), 11553, 2025.
13. Z. Zheng, M. He, X. Yu, J. Li, **L. Chen**, ..., J. Zhang, Y. Zhou, R. Luo. Accelerated long-read variant calling with Clair3 for whole-genome sequencing. *Bioinformatics (IF: 5.50)*, btac181, 2026.
14. X. Yu, Z. Zheng, **L. Chen**, Z. Qin, M. He, R. Luo. Leveraging ONT move table values for signal aware variant calling. *ECCB & Bioinformatics (IF: 5.50)*. 2026.
15. Z. Zheng, M. He, X. Yu, **L. Chen**, J. Li, J. Zhang, R. Luo. Telomere-to-telomere CHM13 reference reveals missing truth variants and improves deep learning-based variant calling in long-read sequencing data. *Quantitative Biology (IF: 2.20)*. 2026.
16. Z. Zheng, Y. Ren, **L. Chen**, A. Wong, S. Li, X. Yu, T. Lam, R. Luo. Repun: an accurate small variant representation unification method for multiple sequencing platforms. *Briefings in Bioinformatics (IF: 7.30)*, 26(1), bbae613, 2025.
17. J. Zhang, Y. Zhou, Y. Ren, M. Au, K. Chow, **L. Chen**, Y. Zhao, J. Su, R. Luo. Toward owner governance in genomic data privacy with Governome. *Cell Reports Methods (IF: 5.80)*, 5(9), 2025.
18. S. Li, F. Chow, **L. Chen**, J. Su, S. Ahmed, R. Luo. MirtronStructDB: A Comprehensive Database of Mirtrons with Predicted Secondary Structures. *F1000Research*, 13, 899, 2024.
19. X. Zhang, **L. Chen**. MulMarker: a comprehensive framework for identifying multi-gene prognostic signatures. *BIBM (CCF B)* (pp. 4014-4021). IEEE. 2023.
20. G. Yuan, F. Yuan, Y. Li, B. Kong, S. Li, **L. Chen**, ..., Z. Li, Y. Xu, X. Qie. Tenrec: A large-scale multipurpose benchmark dataset for recommender systems. *NeurIPS (CCF A)*, 35, 11480-11493, 2022.
21. Q. Jiang, **L. Chen**, W. Zhao, M. Yang. Toward aspect-level sentiment modification without parallel data. *IEEE Intelligent Systems (IF: 4.50)*, 36(1), 75-81, 2021.

22. M. Yang, **L. Chen**, Z. Lyu, J. Liu, Y. Shen, Q. Wu. Hierarchical fusion of common sense knowledge and classifier decisions for answer selection in community question answering. *Neural Networks (IF: 7.20)*, 132, 53-65, 2020.
23. Q. Jiang, **L. Chen**, R. Xu, X. Ao, M. Yang. A challenge dataset and effective models for aspect-based sentiment analysis. *EMNLP (CCF B)* (pp. 6280-6285), 2019.
24. M. Yang, W. Zhou, **L. Chen**, Q. Qu, Z. Zhao, Y. Shen. Investigating the transferring capability of capsule networks for text classification. *Neural Networks (IF: 7.20)*, 118, 247-261, 2019.
25. M. Yang, **L. Chen**, X. Chen, Q. Wu, W. Zhou, Y. Shen. Knowledge-enhanced Hierarchical Attention for Community Question Answering with Multi-task and Adaptive Learning. *IJCAI (CCF B)* (pp. 5349-5355), 2019.
26. M. Yang, J. Liu, **L. Chen**, Z. Zhao, X. Chen, Y. Shen. An advanced deep generative framework for temporal link prediction in dynamic networks. *IEEE Transactions on Cybernetics (IF: 11.30)*, 50(12), 4946-4957, 2019.

Granted patents

1. IMPLEMENTATION METHOD, DEVICE AND EQUIPMENT OF ADAPTIVE PROTEIN PREDICTION FRAMEWORK. (2021)
 - CN patent no. CN202110600871.7. Lei Chen, Min Yang, ..., Qingshan Jiang.
2. A KNOWLEDGE DISTILLATION METHOD FOR SEQUENTIAL RECOMMENDATION BASED ON EARTH MOVER'S DISTANCE. (2020)
 - CN patent no. CN202011245696.6. Lei Chen, Min Yang, ..., Qingshan Jiang.
3. A METHOD FOR CONSTRUCTING SEQUENTIAL RECOMMENDATION MODEL AND SEQUENTIAL RECOMMENDATION METHOD. (2020)
 - CN patent no. CN202010844315.X. Lei Chen, Min Yang, ..., Qingshan Jiang.
4. A SEQUENTIAL RECOMMENDATION METHOD AND SYSTEM BASED ON ADAPTIVE NETWORK DEPTH. (2020)
 - CN patent no. CN202010835626.X. Lei Chen, Min Yang, ..., Qingshan Jiang.
5. AN ADAPTIVE OUTPUT SEQUENTIAL RECOMMENDATION METHOD AND SYSTEM. (2020)
 - CN patent no. CN202010830579.X. Lei Chen, Min Yang, ..., Qingshan Jiang.
6. A SEQUENTIAL RECOMMENDATION METHOD AND SYSTEM BASED ON IMPROVED RESNET STRUCTURE. (2020)
 - CN patent no. CN202010816525.8. Lei Chen, Min Yang, ..., Qingshan Jiang.
7. A CROSS-DOMAIN SEQUENTIAL RECOMMENDATION METHOD BASED ON ADAPTIVE FINE-TUNING STRATEGY. (2019)

-
- CN patent no. CN201911272877.5. Lei Chen, Min Yang, ..., Chengming Li.

Academic services

- Head teaching assistant for *COMP2501 Introduction to Data Science, HKU*. 2022-2026.
- Reviewer for journals and conferences including *Nature Communications, Neurocomputing, Knowledge-based Systems, NeurIPS, ACL, EMNLP, AAAI, COLING, ACM BCB*, etc.
- Organizer of *NLPCC-2020 Shared Task: Multi-Aspect-based Multi-Sentiment Analysis (MAMS)*.
- Volunteer of *GIW XXXIV ISCB-Asia 2025, AAAI-2021*, etc.