

Tiejin Chen

📞 734-210-2624 ✉ tiejin@asu.edu 🌐 [Website](#) 🎓 [Google Scholar](#)

Education

Arizona State University, Tempe <i>PhD in Computer Science</i> • Advisor: Hua Wei; GPA: 3.91/4.00 • Research Interests: Trustworthy LLMs, including uncertainty quantification, safety, explanation, and their applications, especially improving post-training like Supervised Fine-tuning and Reinforcement Learning.	2023-2027 (Expected) Tempe, AZ
University of Michigan, Ann Arbor <i>MS in Applied Statistics</i> • GPA: 3.98/4.00	2021-2023 Ann Arbor, MI
Sichuan University <i>Bachelor in Statistics</i> • GPA: 3.55/4.00 86.59/100; Rank: 4/33	2016-2020 Chengdu, Sichuan

Experience

Amazon <i>Applied Scientist Intern</i> • Working on a project that improves the performance of using LLMs to simulate human behavior;	05/2026 – 08/2026 Sunnyvale
Amazon <i>Applied Scientist Intern</i> • Post-training on small LLM on various tasks to replace Claude's Model in the production pipeline; • Reduce the overall cost to 1%-5% of the original cost with comparable performance.	08/2025 – 11/2025 Sunnyvale
Toyota InfoTech Lab <i>STEM Coop</i> • Build the multi-agent systems that generate the personalized trip plan based on the user's requirement; • Research about the orchestration of multi-agent systems and write a patent about how to detect similar agents inside orchestration and remove the ambiguity provided by the similar agents; • Involved in 4 patents about the hybrid autoregressive language models and diffusion language model, continue learning for LLMs, and increase the performance of orchestration of multi-agent systems. • The submitted paper is accepted by IV 2026.	05/2025 – 08/2025 Mountain View

Selected Publications

[NeurIPS'26]	Tiejin Chen , Maunil Rohitbhai Vyas, Huaiyuan Yao, Hua Wei. <i>Steering Vectors as a Training Signal in LLM Post-Training</i> . The Fortieth Annual Conference on Neural Information Processing Systems
[SDM'26]	Het Patel, Tiejin Chen , Hua Wei, Evangelos E Papalexakis, Jia Chen. <i>Are LLM Uncertainty and Correctness Encoded by the Same Features? A Functional Dissociation via Sparse Autoencoders</i> . SIAM International Conference on Data Mining 2026
[KDD'26 - Blue Sky]	Tiejin Chen* , Xiaoou Liu*, Weibo Li, Xiyang Hu, Hua Wei. <i>The Sim-to-Real Gap of Foundation Model Agents: A Unified MDP Perspective</i> . In Proceedings of the 32nd ACM SIGKDD international conference on Knowledge discovery and data mining - Blue Sky Track
[ICML'26]	Tiejin Chen , Longchao Da, Xiaoou Liu, Hua Wei. <i>Position: Uncertainty Quantification in LLMs is Just Unsupervised Clustering</i> . Forty-Third International Conference on Machine Learning - Position Track
[ICML'26]	Xiaoou Liu, Tiejin Chen , Dengjia Zhang, Yaqing Wang, Lu Cheng, Hua Wei. <i>Diagnosing Multi-step Reasoning Failures in Black-box LLMs via Stepwise Confidence Attribution</i> . Forty-Third International Conference on Machine Learning

[ACL'26]	Tiejun Chen , Huaiyuan Yao, Jia Chen, Evangelos E. Papalexakis, Hua Wei. <i>Every Response Counts: Quantifying Uncertainty of LLM-based Multi-Agent Systems through Tensor Decomposition</i> . The 64th Annual Meeting of the Association for Computational Linguistics
[TMLR]	Yuanpu Cao*, Tiejun Chen *, Yujia Wang*, Cho-Jui Hsieh, Jinghui Chen. <i>Federated Learning with Projected Trajectory Regularization</i> . Transactions on Machine Learning Research
[ACL Findings'26]	Tiejun Chen , Xiaoou Liu, Longchao Da, Jia Chen, Evangelos E. Papalexakis, Hua Wei. <i>Uncertainty Quantification of Large Language Models through Multi-Dimensional Responses</i> . Findings of The 64th Annual Meeting of the Association for Computational Linguistics
[IV'26]	Tiejun Chen , Ahmadreza Moradipari, Kyungtae Han, Hua Wei, Nejib Ammar. <i>Agentic AI for Trip Planning Optimization Application</i> . IEEE Intelligent Vehicles Symposium 2026
[EACL'26 - Oral]	Tiejun Chen *, Kaishen Wang*, Hua Wei. <i>Zer0-Jack: A Memory-efficient Gradient-based Jailbreaking Method for Black-box Multi-modal Large Language Models</i> . The 19th Conference of the European Chapter of the Association for Computational Linguistics; also presented at GenSafeAI Workshop @ NeurIPS 2024
[EACL Findings'26]	Tiejun Chen , Xiaoou Liu, Vishnu Nandam, Kuan-Ru Liou, Hua Wei. <i>Conformal Feedback Alignment: Quantifying Answer-Level Reliability for Robust LLM Alignment</i> . Findings of The 19th Conference of the European Chapter of the Association for Computational Linguistics
[ACL Findings'25]	Tiejun Chen , Pingzhi Li, Kaixiong Zhou, Tianlong Chen, Hua Wei. <i>Vision Language Model Helps Private Information De-Identification in Vision Data</i> . Findings of the 63rd Annual Meeting of the Association for Computational Linguistics
[KDD'25]	Tiejun Chen *, Xiaoou Liu*, Longchao Da, Chacha Chen, Zhen Lin, Hua Wei. <i>Uncertainty Quantification and Confidence Calibration in Large Language Models: A Survey</i> . Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining
[SDM'25]	Tiejun Chen , Longchao Da, Huixue Zhou, Pingzhi Li, Kaizhong Zhou, Tianlong Chen, Hua Wei. <i>Protecting Privacy against Membership Inference Attack with LLM Fine-tuning through Flatness</i> . SIAM International Conference on Data Mining 2025; short version presented at SeTLLM Workshop @ ICLR 2024
[AAAI'24]	Kai Ye, Tiejun Chen , Hua Wei, Liang Zhan. <i>Uncertainty Regularized Evidential Regression</i> . Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence

Service

Teaching Assistant: CSE 355, Intro Theoretical Computer Sci, Spring 2025. CSE 572, Data Mining, Spring 2026.
Reviewer: PAKDD 2024, ICLR 2024 SeTLLM Workshop, WACV 2025, Neurips 2024 SafeGenAI Workshop, ICLR 2025, SDM 2025, CVPR2025, ICCV 2025, ACL 2025 Demo Track, EACL 2026, ICLR 2026, KDD 2026.

Award

Sichuan University Comprehensive Scholarship
ASU Experiential Learning Grant
ASU Graduate Student Government Travel Grant
NSF travel award
Ferdinand A. Stanchi Fellowship

Technical Skills

Technologies: Python, OpenRLHF, Amazon AWS, PyTorch, Transformers, Large Language Models, Reinforcement Learning, Large Language Model Fine-tuning, Multi-agent System (AutoGen).
Hobbies: Mystery Novels, Oscar Prediction
Extracurricular Activities: Deputy director of Reasoning Association for organizing mystery games and organizing Sichuan University to join in the national BBS mystery contest.

Full Publication List

- [NeurIPS'26] **Tiejun Chen**, Maunil Rohitbhai Vyas, Huaiyuan Yao, Hua Wei. *Steering Vectors as a Training Signal in LLM Post-Training*. The Fortieth Annual Conference on Neural Information Processing Systems
- [AAACL-IJCNLP'26] Xiaou Liu, **Tiejun Chen**, Hanyang Chen, Yuan Ling, Hua Wei. *Diverse Reasoning Paths Matter: Symbolic Equivalence Augmentation for Enhancing RL Exploration*. 15th International Joint Conference on Natural Language Processing and the 5th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics
- [AAACL-IJCNLP'26] **Tiejun Chen***, Xiaou Liu*, Ahmadreza Moradipari, Michael Clifford, Kyungtae Han, Yitao Chen, Nejib Ammar, Hua Wei. *Detect Before You Disambiguate: Perturbation-Based Input Ambiguity Detection in LLMs*. 15th International Joint Conference on Natural Language Processing and the 5th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics
- [SDM'26] Het Patel, **Tiejun Chen**, Hua Wei, Evangelos E Papalexakis, Jia Chen. *Are LLM Uncertainty and Correctness Encoded by the Same Features? A Functional Dissociation via Sparse Autoencoders*. SIAM International Conference on Data Mining 2026
- [KDD'26 - Blue Sky] **Tiejun Chen***, Xiaou Liu*, Weibo Li, Xiyang Hu, Hua Wei. *The Sim-to-Real Gap of Foundation Model Agents: A Unified MDP Perspective*. In Proceedings of the 32nd ACM SIGKDD international conference on Knowledge discovery and data mining - Blue Sky Track
- [TMLR] Yuanpu Cao*, **Tiejun Chen***, Yujia Wang*, Cho-Jui Hsieh, Jinghui Chen. *Federated Learning with Projected Trajectory Regularization*. Transactions on Machine Learning Research
- [ICML'26] **Tiejun Chen**, Longchao Da, Xiaou Liu, Hua Wei. *Position: Uncertainty Quantification in LLMs is Just Unsupervised Clustering*. Forty-Third International Conference on Machine Learning - Position Track
- [ICML'26] Xiaou Liu, **Tiejun Chen**, Dengjia Zhang, Yaqing Wang, Lu Cheng, Hua Wei. *Diagnosing Multi-step Reasoning Failures in Black-box LLMs via Stepwise Confidence Attribution*. Forty-Third International Conference on Machine Learning
- [ACL'26] **Tiejun Chen**, Huaiyuan Yao, Jia Chen, Evangelos E. Papalexakis, Hua Wei. *Every Response Counts: Quantifying Uncertainty of LLM-based Multi-Agent Systems through Tensor Decomposition*. The 64th Annual Meeting of the Association for Computational Linguistics
- [ACL Findings'26] **Tiejun Chen**, Xiaou Liu, Longchao Da, Jia Chen, Evangelos E. Papalexakis, Hua Wei. *Uncertainty Quantification of Large Language Models through Multi-Dimensional Responses*. Findings of The 64th Annual Meeting of the Association for Computational Linguistics
- [IV'26] **Tiejun Chen**, Ahmadreza Moradipari, Kyungtae Han, Hua Wei, Nejib Ammar. *Agentic AI for Trip Planning Optimization Application*. IEEE Intelligent Vehicles Symposium 2026
- [EAACL'26 - Oral] **Tiejun Chen***, Kaishen Wang*, Hua Wei. *Zer0-Jack: A Memory-efficient Gradient-based Jailbreaking Method for Black-box Multi-modal Large Language Models*. The 19th Conference of the European Chapter of the Association for Computational Linguistics; also presented at GenSafeAI Workshop @ NeurIPS 2024
- [EAACL Findings'26] **Tiejun Chen**, Xiaou Liu, Vishnu Nandam, Kuan-Ru Liou, Hua Wei. *Conformal Feedback Alignment: Quantifying Answer-Level Reliability for Robust LLM Alignment*. Findings of The 19th Conference of the European Chapter of the Association for Computational Linguistics
- [KDD Explor.] **Tiejun Chen**, Wenwang Huang, Linsey Pang, Dongsheng Luo, Hua Wei. *Are Classification Robustness and Explanation Robustness Really Strongly Correlated? An Analysis Through Input Loss Landscape*. ACM SIGKDD Explorations Newsletter(IF=4.0)

- [QCE'25] Ningyi Xie, Xinwei Lee, **Tiejing Chen**, Yoshiyuki Saito, Nobuyoshi Asai, Dongsheng Cai. *An Adaptive Weighted QITE-VQE Algorithm for Combinatorial Optimization Problems*. IEEE International Conference on Quantum Computing and Engineering 2025
- [ICCPs'25] Yiran Zhang, Khoa Vo, Longchao Da, **Tiejing Chen**, Xiaoou Liu, Hua Wei. *Reproducible and Low-cost Sim-to-Real Environment for Traffic Signal Control*. ACM/IEEE 16th International Conference on Cyber-Physical Systems
- [ACL Findings'25] **Tiejing Chen**, Pingzhi Li, Kaixiong Zhou, Tianlong Chen, Hua Wei. *Unveiling Privacy Risks in Multi-modal Large Language Models: Task-specific Vulnerabilities and Mitigation Challenges*. Findings of the 63rd Annual Meeting of the Association for Computational Linguistics
- [ACL Findings'25] **Tiejing Chen**, Pingzhi Li, Kaixiong Zhou, Tianlong Chen, Hua Wei. *Vision Language Model Helps Private Information De-Identification in Vision Data*. Findings of the 63rd Annual Meeting of the Association for Computational Linguistics
- [KDD'25] **Tiejing Chen***, Xiaoou Liu*, Longchao Da, Chacha Chen, Zhen Lin, Hua Wei. *Uncertainty Quantification and Confidence Calibration in Large Language Models: A Survey*. Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining
- [RLC'25] Justin Turnau, Longchao Da, Khoa Vo, Ferdous Al Rafi, Shreyas Bachiraju, **Tiejing Chen**, Hua Wei. *Joint-Local Grounded Action Transformation for Sim-to-Real Transfer in Multi-Agent Traffic Control*. Reinforcement Learning Conference 2025
- [SDM'25] **Tiejing Chen**, Longchao Da, Huixue Zhou, Pingzhi Li, Kaizhong Zhou, Tianlong Chen, Hua Wei. *Protecting Privacy against Membership Inference Attack with LLM Fine-tuning through Flatness*. SIAM International Conference on Data Mining 2025; short version presented at SeTLLM Workshop @ ICLR 2024
- [ICLR Workshop'25] **Tiejing Chen***, Kuan-Ru Liou*, Mithun Shivakoti, Aaryan Gaur, Pragya Kumari, Meiqi Guo, Hua Wei. *Abg-SciQA: A Dataset for Understanding and Resolving Ambiguity in Scientific Questions*. ICLR 2025 Workshop on Navigating and Addressing Data Problems for Foundation Models
- [ITSC'24] **Tiejing Chen***, Prithvi Parag Shirke*, Bharatesh Chakravarthi, Arpitsinh Vaghela, Longchao Da, Duo Lu, Yezhou Yang, Hua Wei. *SynTraC: A Synthetic Dataset for Traffic Signal Control from Traffic Monitoring Cameras*. Proceedings of the 27th IEEE International Conference on Intelligent Transportation Systems
- [Phys. Rev. A] Ningyi Xie, Jiahua Xu, **Tiejing Chen**, Yoshiyuki Saito, Nobuyoshi Asai, Dongsheng Cai. *Performance Upper Bound of the Grover-Mixer Quantum Alternating Operator Ansatz*. Physical Review A
- [BlackHat'24] Zicheng Wang, **Tiejing Chen**, Qinrun Dai, Yueqi Chen, Hua Wei, Qingkai Zeng. *When eBPF Meets Machine Learning: On-the-fly OS Kernel Compartmentalization*. BlackHat USA 2024 Briefing
- [IJMLC] Longchao Da, Kuan-Ru Liou, **Tiejing Chen**, Xuesong Zhou, Xiangyong Luo, Yezhou Yang, Hua Wei. *Open-TI: Open Traffic Intelligence with Augmented Language Model*. International Journal of Machine Learning and Cybernetics (IF=5.6)
- [AAAI'24] Kai Ye, **Tiejing Chen**, Hua Wei, Liang Zhan. *Uncertainty Regularized Evidential Regression*. Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence
- [arXiv'25] Longchao Da, **Tiejing Chen**, Zhuoheng Li, Shreyas Bachiraju, Huaiyuan Yao, Xiyang Hu, Zhengzhong Tu, Yue Zhao, Dongjie Wang, Ram Pendyala, Benjamin Stabler, Yezhou Yang, Xuesong Zhou, Hua Wei. *Generative AI in Transportation Planning: A Survey*. arXiv:2503.07158
- [arXiv'24] Longchao Da, **Tiejing Chen**, Lu Cheng, Hua Wei. *LLM Uncertainty Quantification through Directional Entailment Graph and Claim Level Response Augmentation*. arXiv:2407.00994

[arXiv'23]

Tiejun Chen*, Yuanpu Cao*, Yujia Wang*, Cho-Jui Hsieh, Jinghui Chen. *Federated Learning with Projected Trajectory Regularization*. arXiv:2312.14380

[arXiv'22]

Tiejun Chen*, Yicheng Tao*. *Learning Sparsity and Randomness for Data-driven Low Rank Approximation*. arXiv:2212.08186