Jiaao Chen

chenjiaao1998@gmail.com | +1 650-898-9905 | Google Scholar

Summary

AI/ML research scientist and technical leader with 5,000+ Google Scholar citations (h-index: 24, i10-index: 30). Expert in building production-grade agentic AI systems, LLM post-training pipelines (SFT, GRPO/RL), and multi-agent platforms. First author of 20+ publications at top venues (NeurIPS, ICLR, EMNLP, ACL). Proven track record from founding startup role through Meta and Amazon, specializing in LLM reasoning, tool-use agents, and data-efficient training.

Professional Experience

Eigen AI — Palo Alto, USA
Founding Member of Technical Staff — Jan. 2026 – Present

- Own and lead **EigenData** — an integrated, self-evolving multi-agent platform for function-calling data that automates the full lifecycle from environment construction to trajectory synthesis and quality assurance. Architected the three-subsystem design: **DatabaseAgent** for domain database construction, **CodingAgent** for verified executable environment generation with iterative test-debug loops, and **DataAgent** for multi-turn trajectory synthesis with self-evolving prompt optimization.
- Build and maintain **EigenLoop** — the full post-training pipeline (SFT, GRPO/RL, evaluation) powering production-grade agentic AI systems across finance, legal, and enterprise domains.
- Applied EigenData to audit and repair the Berkeley Function-Calling Leaderboard (BFCL-V3), uncovering systematic errors in **71.5%** of entries across function schemas, implementations, and trajectories. Designed outcome-aware evaluation protocols that assess task success via database-state correctness, producing model rankings that substantially better correlate with human judgments. On τ^2-bench, combined self-evolving synthetic data with verifier-based GRPO to achieve **73.0%** pass[1] on Airline and **98.3%** on Telecom, matching or exceeding frontier models using only open-weight backbones.

Meta — Menlo Park, USA
Research Scientist — Mar. 2025 – Jan. 2026

- GenAI team: Led the effort that improved the BFCL function-calling performance of Llama 4 Maverick from ∼50 to **72**, and boosted the internal evaluation benchmark (ToolBench) by **30 percentage points**, through end-to-end agentic post-training from synthetic data generation to reinforcement learning.
- TBD Agent Workstream: Built the agentic synthetic data generation pipeline with verified tasks and LLM-as-judge for agentic RL. Contributed to the RL infrastructure and designed reward functions for mid-training and post-training of TBD Agents, improving multi-turn tool-use and planning capabilities.

Amazon — Palo Alto, USA
Applied Scientist — May 2024 – Mar. 2025

- Improved the reasoning abilities of LLMs for modeling customer shopping behaviors, enabling more accurate intent prediction and personalized recommendation through chain-of-thought and multi-step inference.
- Pre-trained a shopping behavior LLM on large-scale e-commerce interaction data, capturing complex sequential purchasing patterns and user preferences.

Tencent AI (US) — Seattle, USA
Research Intern — May 2023 – Aug. 2023

- Developed Skills-in-Context Prompting, a method that unlocks compositional reasoning in LLMs by teaching models to combine foundational skills to solve complex, unseen problems. Published at EMNLP 2024.

Amazon (AWS) — Santa Clara, USA
Applied Scientist Intern — June 2022 – Jan. 2023

- Designed a unified parameter-efficient fine-tuning framework that systematically explores design spaces across adapter architectures; published at ICLR 2023.
- Developed a cheaper and better diffusion language model with soft-masked noise; published at EMNLP 2023.

Allen Institute for AI (AI2) — Seattle, USA
Research Intern, Mosaic Team — June 2021 – Sep. 2021

- Developed compositional data augmentation methods for commonsense question answering with Yejin Choi's group.

Microsoft Research Seattle, USA
Researcher, Dynamics 365 AI Dec. 2020 – May 2021

- Created HiddenCut, a data augmentation method for NLU with better generalization; published at ACL 2021.
- Proposed Mixture of Virtual Prompts (MVP), an efficient fine-tuning method that learns virtual prompts with distinct capacities and dynamically aggregates them per input, achieving comparable performance to full fine-tuning with only 0.03% of parameters.

Georgia Institute of Technology / Stanford University Atlanta / Stanford, USA
Research Assistant, Advisor: Diyi Yang Aug. 2019 – May 2024

- Published 20+ first-author papers on data-efficient NLP, LLM reasoning, dialogue summarization, and diffusion language models at venues including NeurIPS, ICLR, EMNLP, ACL, and NAACL.

Education

Georgia Institute of Technology / Stanford University Atlanta / Stanford, USA
Ph.D. in Computer Science Aug. 2019 – May 2024
Advisor: Diyi Yang

Zhejiang University Hangzhou, China
B.Eng. in Computer Science and Technology Sep. 2015 – Jul. 2019
Minor in Advanced Honor Class of Engineering Education, Chu Kochen Honors College (40/6,000)

Publications

Google Scholar Citations: 5,032 | *h-index: 24* | *i10-index: 30*

Preprints

1. "EigenData: A Self-Evolving Multi-Agent Platform for Function-Calling Data Synthesis, Auditing, and Repair",
Jiaao Chen, Jingyuan Qi, Mingye Gao, Wei-Chen Wang, Hanrui Wang, and Di Jin
arXiv:2603.05553, 2026
2. "From Self-Evolving Synthetic Data to Verifiable-Reward RL: Post-Training Multi-turn Interactive Tool-Using Agents",
Jiaxuan Gao, **Jiaao Chen**, Chuyi He, Wei-Chen Wang, Shusheng Xu, Hanrui Wang, Di Jin, and Yi Wu
arXiv:2601.22607, 2026
3. "Maslab: A Unified and Comprehensive Codebase for LLM-based Multi-Agent Systems",
Rui Ye, Keduan Huang, Qimin Wu, Yuzhu Cai, Tian Jin, Xianghe Pang, Xiangrui Liu, Jiaqi Su, Chen Qian, Bohan Tang, et al
arXiv:2505.16988, 2025
4. "Agentic Workflows for Conversational Human-AI Interaction Design",
Arthur Caetano, Kavya Verma, Atieh Taheri, Radha Kumaran, Zichen Chen, **Jiaao Chen**, Tobias Höllerer, and Misha Sra
arXiv:2501.18002, 2025
5. "Standard Benchmarks Fail – Auditing LLM Agents in Finance Must Prioritize Risk",
Zichen Chen, **Jiaao Chen**, Jianda Chen, and Misha Sra
arXiv:2502.15865, 2025
6. "Are We There Yet? Revealing the Risks of Utilizing Large Language Models in Scholarly Peer Review",
Rui Ye, Xianghe Pang, Jingyi Chai, **Jiaao Chen**, Zhenfei Yin, Zhen Xiang, Xiaowen Dong, Jing Shao, and Siheng Chen
arXiv:2412.01708, 2024
7. "Dynamic Skill Adaptation for Large Language Models",
Jiaao Chen and Diyi Yang
arXiv:2412.19361, 2024

Conference & Journal

1. "WorkForceAgent-R1: Incentivizing Reasoning Capability in LLM-based Web Agents via Reinforcement Learning",
Yuchen Zhuang, Di Jin, **Jiaao Chen**, Wenqi Shi, Hanrui Wang, and Chao Zhang
European Chapter of the Association for Computational Linguistics (EACL), 2026
2. "Learning Multi-step Reasoning via Persistent Latent State Propagation",

Yixiao Li, **Jiaao Chen**, Xin Tang, Fangzhou Wu, Jiahui Yu, Haotian Qi, Wei Xuan, Hao Zhao, Peng Nie, and Di Jin
Workshop on Latent & Implicit Thinking — Going Beyond CoT, 2026

3. "From Tasks to Teams: A Risk-First Evaluation Framework for Multi-Agent LLM Systems in Finance",
Zichen Chen, Jianda Chen, **Jiaao Chen**, and Misha Sra
ICML 2025 Workshop on Reliable and Responsible Foundation Models

4. "DARG: Dynamic Evaluation of Large Language Models via Adaptive Reasoning Graph",
Zhehao Zhang, **Jiaao Chen**, and Diyi Yang
Conference on Neural Information Processing Systems (NeurIPS), 2024

5. "From Scroll to Misbelief: Modeling the Unobservable Susceptibility to Misinformation on Social Media",
Yanchen Liu, Mingyu Derek Ma, Wenna Qin, Azure Zhou, **Jiaao Chen**, Weiyan Shi, Wei Wang and Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024

6. "Skills-in-context Prompting: Unlocking Compositionality in Large Language Models",
Jiaao Chen, Xiaoman Pan, Dian Yu, Kaiqiang Song, Xiaoyang Wang, Dong Yu and Jianshu Chen
Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP Findings), 2024

7. "DyVal: Graph-informed Dynamic Evaluation of Large Language Models",
Kaijie Zhu*, **Jiaao Chen***, Jindong Wang, Neil Zhenqiang Gong, Diyi Yang, Xing Xie
International Conference on Learning Representations (ICLR), 2024

8. "Can Large Language Models Transform Computational Social Science?",
Caleb Ziems, William Held, Omar Shaikh, **Jiaao Chen**, Zhehao Zhang and Diyi Yang
Computational Linguistics, 2024

9. "Unlearn What You Want to Forget: Efficient Unlearning for LLMs",
Jiaao Chen and Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023

10. "A Cheaper and Better Diffusion Language Model with Soft-Masked Noise",
Jiaao Chen and Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023

11. "Is ChatGPT a General-Purpose Natural Language Processing Task Solver?",
Chengwei Qin, Aston Zhang, Zhuosheng Zhang, **Jiaao Chen**, Michihiro Yasunaga, and Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023

12. "Mitigating Biases in Hate Speech Detection from A Causal Perspective",
Zhehao Zhang, **Jiaao Chen**, and Diyi Yang
Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP Findings), 2023

13. "Where Does Your News Come From? Predicting Information Pathways in Social Media",
Alexander K Taylor, Nuan Wen, Po-Nien Kung, **Jiaao Chen**, Violet Peng, and Wei Wang
ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2023

14. "Controllable Conversation Generation with Conversation Structures via Diffusion Models",
Jiaao Chen, and Diyi Yang
Findings of the Annual Meeting of the Association for Computational Linguistics (ACL Findings), 2023

15. "Human-in-the-loop Abstractive Dialogue Summarization",
Jiaao Chen, Mohan Dodda, and Diyi Yang
Findings of the Annual Meeting of the Association for Computational Linguistics (ACL Findings), 2023

16. "Compositional Data Augmentation for Abstractive Conversation Summarization",
Siru Ouyang, **Jiaao Chen**, Jiawei Han, and Diyi Yang
Annual Meeting of the Association for Computational Linguistics (ACL), 2023

17. "Parameter-Efficient Fine-Tuning Design Spaces",
Jiaao Chen, Aston Zhang, Xingjian Shi, Mu Li, Alex Smola, and Diyi Yang
International Conference on Learning Representations (ICLR), 2023

18. "An Empirical Survey of Data Augmentation for Limited Data Learning in NLP",
Jiaao Chen*, Derek Tam*, Colin Raffel, Mohit Bansal, and Diyi Yang
Transactions of the Association for Computational Linguistics (TACL)

19. "When FLUE Meets FLANG: Benchmarks and Large Pretrained Language Model for Financial Domain",
Raj Sanjay Shah, Kunal Chawla, Dheeraj Eidnani, Agam Shah, Wendi Du, Sudheer Chava, Natraj Raman, Charese Smiley, **Jiaao Chen**, and Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022

20. "E-VALUE: Understanding Dialect Disparity in NLU",
Caleb Ziems*, **Jiaao Chen***, Camille Harris, Jessica Anderson, and Diyi Yang
Annual Meeting of the Association for Computational Linguistics (ACL), 2022

21. "Focus on the Action: Learning to Highlight and Summarize Jointly for Email To-Do Items Summarization",
Kexun Zhang, **Jiaao Chen**, Diyi Yang
Annual Meeting of the Association for Computational Linguistics (ACL), 2022

22. "Leveraging Expert Guided Adversarial Augmentation For Improving Generalization in Named Entity Recognition",
Aaron Reich, **Jiaao Chen**, Aastha Agrawal, Yanzhe Zhang, Diyi Yang
Findings of the Annual Meeting of the Association for Computational Linguistics (ACL Findings), 2022

23. "Simple Conversational Data Augmentation for Semi-supervised Abstractive Dialogue Summarization",
Jiaao Chen, Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021

24. "HiddenCut: Simple Data Augmentation for Natural Language Understanding with Better Generalization",
Jiaao Chen, Dinghan Shen, Weizhu Chen, and Diyi Yang
Annual Meeting of the Association for Computational Linguistics (ACL), 2021

25. "Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs",
Jiaao Chen, Diyi Yang
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021

26. "Continual Learning for Text Classification with Information Disentanglement Based Regularization",
Yufan Huang, Yanzhe Zhang, **Jiaao Chen**, Xuezhi Wang and Diyi Yang
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021

27. "Weakly-Supervised Hierarchical Models for Predicting Persuasive Strategies in Good-faith Textual Requests",
Jiaao Chen, Diyi Yang
AAAI Conference on Artificial Intelligence (AAAI), 2020

28. "Multi-View Sequence-to-Sequence Models with Conversational Structure for Abstractive Dialogue Summarization",
Jiaao Chen, Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020

29. "Local Additivity Based Data Augmentation for Semi-supervised NER",
Jiaao Chen*, Zhenghui Wang*, Ran Tian, Zichao Yang and Diyi Yang
Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020

30. "Examining the Ordering of Rhetorical Strategies in Persuasive Requests",
Omar Shaikh, **Jiaao Chen**, Jon Saad-Falcon, Polo Chau and Diyi Yang
Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP Findings), 2020

31. "MixText: Linguistically-Informed Interpolation of Hidden Space for Semi-Supervised Text Classification",
Jiaao Chen, Zichao Yang, Diyi Yang
Annual Meeting of the Association for Computational Linguistics (ACL), 2020

32. "Semi-supervised Models via Data Augmentation for Classifying Interactive Affective Responses",
Jiaao Chen*, Yuwei Wu*, Diyi Yang
AAAI Workshop on Affective Content Analysis, 2020

33. "Let's Make Your Request More Persuasive: Modeling Persuasive Strategies via Semi Supervised Neural Nets on Crowdfunding Platforms",
Diyi Yang*, **Jiaao Chen***, Zichao Yang, Dan Jurafsky, Eduard Hovy
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019

34. "Incorporating Structured Commonsense Knowledge in Story Completion",
Jiaao Chen, Jianshu Chen, Zhou Yu
AAAI Conference on Artificial Intelligence (AAAI), 2019

Teaching & Advising

Teaching Assistant, NLP (CS-4650/7650), Georgia Tech, with Prof. Mark Riedl — Spring 2023
Teaching Assistant, NLP (CS-4650/7650), Georgia Tech, with Prof. Diyi Yang — Spring 2020
Teaching Assistant, Computer Organization, Zhejiang University — Spring 2018

Advised Master's Students: Aaron Reich, Aastha Agrawal, Yufan Huang, Kunal Chawla, Nikhil Gupta (2020–2022)
Advised Undergraduate Students: Zhehao Zhang, Siru Ouyang, Mohan Dodda, Wenna Qin, Kexun Zhang, Dheeraj Eidnani, Yanzhe Zhang, Omar Shaikh, Jon Saad-Falcon (2019–2024)
Co-advised students published at ACL, EMNLP, NeurIPS, and SIGIR.

Professional Service

Area Chair: ARR 2025 (April, October)
Program Committee / Reviewer: ACL (2021–2024), EMNLP (2020–2024), NAACL 2021, ICLR (2022–2023), NeurIPS 2025, ICML 2025, CHI 2023, UIST 2022, WWW 2021, AAAI (2020–2021), IJHCI 2023
Conference Volunteer: ACL 2020
Organizer: Georgia Tech NLP Seminar (2019–2021)

Selected Invited Talks

- "Efficient and Adaptive Machine Learning for NLP" — Nvidia, Salesforce, 2024
- "Efficient and Adaptive Machine Learning for NLP" — Zhejiang University, 2024
- "Learning with Limited Data in NLP" — Aalborg University, 2024
- "An Empirical Survey of Data Augmentation for Limited Data Learning in NLP" — Conference Talk, ACL, 2023
- "Human-in-the-loop Abstractive Dialogue Summarization" — Adobe, 2023
- "Simple Conversational Data Augmentation for Semi-supervised Abstractive Dialogue Summarization" — Conference Talk, EMNLP, 2021
- "HiddenCut: Simple Data Augmentation for NLU with Better Generalization" — Conference Talk, ACL, 2021
- "Structure-Aware Abstractive Conversation Summarization via Discourse and Action Graphs" — Conference Talk, NAACL, 2021
- "Conversation Summarization with Multiple Dialogue Views" — Google NLP, 2020
- "Multi-View Sequence-to-Sequence Models with Conversational Structure for Abstractive Dialogue Summarization" — Conference Talk, EMNLP, 2020
- "Local Additivity Based Data Augmentation for Semi-supervised NER" — Conference Talk, EMNLP, 2020
- "MixText: Linguistically-Informed Interpolation of Hidden Space for Semi-Supervised Text Classification" — Conference Talk, ACL, 2020
- "Incorporating Structured Commonsense Knowledge in Story Completion" — Conference Talk, AAAI, 2019

Selected Awards & Honors

- Outstanding Undergraduate Thesis, Zhejiang University — 2019
- Honorable Mention, MCM/ICM Mathematical Contest in Modeling — 2017
- Academic Physics Contest, Zhejiang University, fourth place — 2017
- ACEE Honor Class, Chu Kochen Honors College, Zhejiang University (40/6,000) — 2016
- Scholarship for Academic Achievement, Zhejiang University — 2016
- Scholarship for All-around Achievement, Zhejiang University — 2016
- ACM School Trial, Zhejiang University, third prize — 2016

Technical Skills

Core: Python, PyTorch, CUDA, Distributed Training (DeepSpeed, FSDP), Reinforcement Learning (GRPO, PPO)
LLM/Agent Stack: vLLM, Hugging Face Transformers, LangChain, FastMCP, SFT/RLHF Pipelines
Infrastructure: SQL, Docker, Git, AWS, GCP
Programming: C/C++, TensorFlow, Verilog, Matlab, SQL, C#
Interests: Piano (Level 10 Certificate), Basketball