Junzhi Chen

+1-347-431-8891 | jc13140@nyu.edu | google scholar | homepage

Education

- **The Hong Kong University of Science and Technology, Guangzhou** *Starting Sep 2026*
  Incoming PhD Student - Artificial Intelligence Guangzhou, China
- **New York University** *Sep 2024 - May 2026*
  Master of Science in Computer Engineering New York, United States
- **The Chinese University of Hong Kong, Shenzhen** *Sep 2020 - May 2024*
  Bachelor of Engineering in Computer Science and Engineering Shenzhen, China

Publications

[1] *AppWorld-UL: Benchmarking Diverse Agent-User Interactions for Tool-Use*
Junzhi Chen, Harsh Trivedi, Jane Pan, Michael JQ Zhang, Tejas Srinivasan, Niranjan Balasubramanian, Ashish Sabharwal
ICML 2026

[2] *Smurfs: Multi-Agent System using Context-Efficient DFSDT for Tool Planning*
Junzhi Chen, Juhao Liang, Benyou Wang
NAACL 2025

Research Experience

- **Machine Learning for Language (ML²) Lab, NYU** *Nov 2024 - May 2026*
  Research Assistant New York, United States
  - Designed a perturbation-based method to transform autonomous agent tasks into user-in-the-loop tasks for three interaction types: underspecification, infeasibility, and confirmation-seeking.
  - Led the development of AppWorld-UL, a user-in-the-loop benchmark of 516 manually designed tasks that combine long-horizon tool use with adaptive user communication in a large stateful environment.
  - Evaluated frontier Agents, showing that the best Claude Opus 4.7 ReAct Coding agent achieved only 48.6% I-TGC, highlighting agent-user interaction as a key bottleneck.
- **Shenzhen Research Institute of Big Data, CUHKSZ** *Jun 2023 - Aug 2024*
  Research Assistant Shenzhen, China
  - Identified limitations in the DFSDT (Depth-First Search Decision Tree) agent-planning algorithm.
  - Proposed "Smurfs", a novel multi-agent system that enhances DFSDT with a modular, context-efficient, and training-free design. Project Page
  - Reduced token usage by 60.9% compared to DFSDT and enabled Mistral-7b to perform on par with GPT-4-DFSDT on StableToolBench.

Work Experience

- **ModelBest, Shenzhen Intermediate People's Court** *Mar 2024 - Jul 2024*
  Machine Learning Engineer Intern Shenzhen, China
  - Built large-scale pretraining dataset for legal-domain LLMs, performing filtering, deduplication, and quality selection over 15TB of court documents.
  - Developed an instruction-tuning pipeline by synthesizing high-quality supervised data from the judicial database and iteratively refining models based on feedback from judges.
  - Contributed to the deployment of the first Chinese LLM for Judicial Trials.

Project

- **New York University** *Jan 2025 - May 2025*
  DS-GA 1012: Natural Language Understanding New York, United States
  - Developed a memory-augmented LLM to use explicit memory (sparse attention key-value pairs) for math, code and general reasoning tasks.
  - Implemented the codebase for knowledge base construction, memory writing/reading/sparsification, model inference, training, and evaluation.
  - Detailed design can be seen at project page and project report.