Jiacheng Liu

✉ liujch1998@gmail.com   ✉ liujc@cs.washington.edu   ⌂ liujch1998.github.io   ⌨ liujch1998   🐦 liujc1998

INTERESTS

I love working with **LLM data, pre-training**, and their connections. I want to figure out how to make LLMs with native **memory** mechanism and dynamically evolve. I'm also fascinated with building **frontier AI applications.**

EDUCATION

UNIVERSITY OF WASHINGTON   Seattle, WA, USA

Ph.D. in Computer Science and Engineering. Research area: natural language processing   2021.03 - Present

Advisors: Yejin Choi, Hannaneh Hajishirzi

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN (UIUC)   Urbana, IL, USA

B.S. in Computer Science; Minor in Physics; GPA 3.97/4.0   2016.08 - 2019.12

Bronze Tablet (3%); Degree with Highest Honors; James Scholar; Dean's List: Fa'16 through Sp'19

SELECTED PUBLICATIONS

Infini-gram: Scaling Unbounded n-gram Language Models to a Trillion Tokens

Jiacheng Liu, Sewon Min, Luke Zettlemoyer, Yejin Choi, Hannaneh Hajishirzi

COLM 2024 (**Oral Spotlight, 2%**); <u>served 700 million API calls</u>; cited by **Jurafsky & Martin** textbook

OLMoTrace: Tracing Language Model Outputs Back to Trillions of Training Tokens

Jiacheng Liu, Taylor Blanton, Yanai Elazar, Sewon Min, YenSung Chen, Arnavi Chheda-Kothary, Huy Tran, Byron Bischoff, Eric Marsh, Michael Schmitz, Cassidy Trier, Aaron Sarnat, Jenna James, Jon Borchardt, Bailey Kuehl, Evie Cheng, Karen Farley, Sruthi Sreeram, Taira Anderson, David Albright, Carissa Schoenick, Luca Soldaini, Dirk Groeneveld, Rock Yuren Pang, Pang Wei Koh, Noah A. Smith, Sophie Lebrecht, Yejin Choi, Hannaneh Hajishirzi, Ali Farhadi, Jesse Dodge

ACL 2025 System Demonstrations Track (**Best Demo Paper**); <u>5k+ users and 20k+ hits</u>

Infini-gram mini: Exact n-gram Search at the Internet Scale with FM-Index

Hao Xu, **Jiacheng Liu**, Yejin Choi, Noah A. Smith, Hannaneh Hajishirzi

Arxiv

2 OLMo 2 Furious

Team OLMo, Pete Walsh, Luca Soldaini, Dirk Groeneveld, Kyle Lo, ..., **Jiacheng Liu**, ..., Ali Farhadi, Noah A. Smith, Hannaneh Hajishirzi

COLM 2025

Establishing Task Scaling Laws via Compute-Efficient Model Ladders

Akshita Bhagia*, **Jiacheng Liu***, Alexander Wettig, David Heineman, Oyvind Tafjord, Ananya Harsh Jha, Luca Soldaini, Noah A. Smith, Dirk Groeneveld, Pang Wei Koh, Jesse Dodge, Hannaneh Hajishirzi

COLM 2025

Unpacking DPO and PPO: Disentangling Best Practices for Learning from Preference Feedback

Hamish Ivison, Yizhong Wang, **Jiacheng Liu**, Zeqiu Wu, Valentina Pyatkin, Nathan Lambert, Noah A Smith, Yejin Choi, Hannaneh Hajishirzi

NeurIPS 2024

Don't throw away your value model! Making PPO even better via Value-Guided Monte-Carlo Tree Search decoding

Jiacheng Liu, Andrew Cohen, Ramakanth Pasunuru, Yejin Choi, Hannaneh Hajishirzi, Asli Celikyilmaz
COLM 2024

MathVista: Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts
Pan Lu, Hritik Bansal, Tony Xia, **Jiacheng Liu**, Chunyuan Li, Hannaneh Hajishirzi, Hao Cheng, Kai-Wei Chang, Michel Galley, Jianfeng Gao
ICLR 2024 (**Oral**); NeurIPS 2023 MATH-AI Workshop

Crystal: Introspective Reasoners Reinforced with Self-Feedback
Jiacheng Liu, Ramakanth Pasunuru, Hannaneh Hajishirzi, Yejin Choi, Asli Celikyilmaz
EMNLP 2023 (Main Conference, **Oral**)

Vera: A General-Purpose Plausibility Estimation Model for Commonsense Statements
Jiacheng Liu*, Wenya Wang*, Dianzhuo Wang, Noah A. Smith, Yejin Choi, Hannaneh Hajishirzi
EMNLP 2023 (Main Conference, **Oral**)

NaturalProver: Grounded Mathematical Proof Generation with Language Models
Sean Welleck*, **Jiacheng Liu***, Ximing Lu, Hannaneh Hajishirzi, Yejin Choi
NeurIPS 2022

Generated Knowledge Prompting for Commonsense Reasoning
Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, Hannaneh Hajishirzi
ACL 2022 (Main Conference)

NaturalProofs: Mathematical Theorem Proving in Natural Language
Sean Welleck, **Jiacheng Liu**, Ronan Le Bras, Hannaneh Hajishirzi, Yejin Choi, Kyunghyun Cho
NeurIPS 2021 Datasets and Benchmarks Track (**Oral, 1%**)

AWARDS

ACL 2025 Best Demo Paper		2025.07
Inverse Scaling Challenge	Third Prize	2023.12
Qualcomm Innovation Fellowship (North America)		2023.08
Meta AI Mentorship Program (2023-2024)		2023.08
Meta AI Mentorship Program (2022-2023)		2022.09
2020 CRA Outstanding Undergraduate Researcher Award	Honorable Mention	2019.12
Correlation One Terminal Live: UIUC	Team 1st Place	2019.09
John R. Pasta Outstanding Undergraduate Award		2019.04
ACM-ICPC World Finals	Team 62nd Place	2019.04
ACM-ICPC Neural Network Challenge	Team 2nd Place	2019.04
ACM-ICPC Mid-Central USA Regional Programming Contest	Team 1st Place	2018.11
UI Undergraduate Math Contest	1st Place	2018.02

INDUSTRY EXPERIENCE

Allen Institute for AI	Student Researcher	2024.06 - Present

Conducted research on language model pretraining, scaling laws, model creativity, and training data attribution

FAIR, Meta	Visiting Researcher	2022.10 - 2024.05

Conducted research in language model reasoning, reinforcement learning, and text decoding algorithms

Oculus, Facebook	Software Engineer	2020.02 - 2021.03

Worked with the natural language generation (NLG) team in Facebook Assistant