

# Jiacheng Liu

✉ liujch1998@gmail.com   ✉ liujc@cs.washington.edu   🐱 liujch1998.github.io   🌐 liujch1998   🐦 liujc1998

## INTERESTS

My work focuses on **LLM data and pretraining**.

I'm also interested in enabling **native test-time memory** and **dynamic evolution** for LLMs.

## 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%**); served 700 million API calls; 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**); 5k+ users and 20k+ hits

### [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**<sup>\*</sup>, Wenya Wang<sup>\*</sup>, Dianzhuo Wang, Noah A. Smith, Yejin Choi, Hannaneh Hajishirzi  
EMNLP 2023 (Main Conference, **Oral**)

**NaturalProver: Grounded Mathematical Proof Generation with Language Models**

Sean Welleck<sup>\*</sup>, **Jiacheng Liu**<sup>\*</sup>, 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

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

## INDUSTRY EXPERIENCE

|   |                     |                   |
|---|---------------------|-------------------|
| <b>Allen Institute for AI</b>   | Student Researcher  | 2024.06 - Present |
| Conducted research on language model pretraining, scaling laws, model creativity, and training data attribution |                     |                   |
| <b>FAIR, Meta</b>   | Visiting Researcher | 2022.10 - 2024.05 |
| Conducted research in language model reasoning, reinforcement learning, and text decoding algorithms            |                     |                   |
| <b>Oculus, Facebook</b>   | Software Engineer   | 2020.02 - 2021.03 |
| Worked with the natural language generation (NLG) team in Facebook Assistant                                    |                     |                   |