Jie Cao

Assistant Professor
(https://www.ou.edu/coe/cs/people/faculty/cao)
School of Computer Science
(http://www.ou.edu/coe/cs)
University of Oklahoma
(https://www.ou.edu/) 110 W. Boyd St., DEH 210-B, Norman, OK 73019

Dr. Jie Cao is an Assistant Professor in the **School of Computer Science (https://www.ou.edu/coe/cs)** at the University of Oklahoma, leading the OUNLP lab. I am also affiliated with the **Data Science and Analytics Institute (https://www.ou.edu/coe/dsai)** at OU. Before joining OU, he spent two years as a post-doctoral researcher at the NSF AI Institute for Student-AI Teaming (**iSAT (https://www.col-orado.edu/research/ai-institute/)**) at the University of Colorado Boulder, where he mainly worked with **Dr. James Martin (https://home.cs.colorado.edu/~martin/)** and **Dr. Martha Palmer (https://www.colorado.edu/faculty/palmer-martha/)**. He obtained his Ph.D. from the **Kahlert School of Computing (https://cs.utah.edu/)** at the **University of Utah (https://www.utah.edu)**, where he worked with **Dr. Vivek Srikumar (https://svivek.com)**. Earlier in his academic journey, he completed his M.S. and B.S. in Computer Science at Huazhong University of Science of Technology~(**HUST (https://english.hust.edu.cn/)**) in China, and he has also worked/interned in industrial companies including Alibaba, Baidu, Sohu, WeChat(@Palo Alto), and Amazon, etc.

Research Interests

I work on **Natural Language Processing** and **Machine Learning**. Current research interests include:

- Multi-party Multi-modal Dialogue/Discourse Analysis on Mental Health, Education, etc
- LLM/MLLM Alignement and Agents, focusing on Domain Specific "World" Model and Human-AI Teaming, etc
- Efficient Structured Prediction and Symbolic Methods for Controlling and Augmenting Neural Networks
- Robust Deployment, and Evaluation of Trustworthy AI

News

- 11/2025: New preprint "Translation via Annotation: A Computational Study of Translating Classical Chinese into Japanese" is now available on **arXiv (https://arxiv.org/abs/2511.05239)**.
- 10/2025: New preprint "Rethinking On-policy Optimization for Query Augmentation" is now available on **arXiv (https://arxiv.org/abs/2510.17139)**. We present the first systematic comparison of prompting-based and RL-based query augmentation and introduce a novel hybrid method OPQE that outperforms both approaches.
- 10/2025: Congratulations to Cuong Huynh! His undergraduate summer research on **text**

Selected Publications

(See full list in **Publication Page (/publications/)** or **Google Scholar (https://scholar.google.com/citations?user=McBrjX4AAAAJ&hl=en)**)

ARXIV	Zilong Li and **Jie Cao**. 2025. **Translation via Annotation: A Computational Study of Translating Classical Chinese into Japanese**. **BibTeX** | **URL (https://arxiv.org/abs/2511.05239)**
TSAR	Cuong Huynh and **Jie Cao**. 2025. **OUNLP at TSAR 2025 Shared Task Multi-Round Text Simplifier via Code Generation**. In Matthew Shardlow, Fernando Alva-Manchego, Kai North, Regina Stodden, Horacio Saggion, Nouran Khallaf, and Akio Hayakawa, editors, *Proceedings of the Fourth Workshop on Text Simplification, Accessibility and Readability (TSAR 2025)*, pages 223–230, Suzhou, China, November. Association for Computational Linguistics. **BibTeX** | **PDF (https://mlciv.com/papers/huynh-cao-2025-ounlp.pdf)** | **URL (https://aclanthology.org/2025.tsar-1.19/)**
ARXIV	Zhichao Xu, Shengyao Zhuang, Xueguang Ma, Bingsen Chen, Yijun Tian, Fengran Mo, **Jie Cao**, and Vivek Srikumar. 2025. **Rethinking On-policy Optimization for Query Augmentation**. *arXiv preprint arXiv:2510.17139*, October. **BibTeX** | **URL (https://arxiv.org/abs/2510.17139)**
J. HYDROL.	Jiaorui Zhang, Haowen Yue, Milad Basirifard, **Jie Cao**, and Tiantian Yang. 2025. **A Mamba-type of Deep State Space Model for Reservoir Release Simulation with a Large-Scale Verification over 441 Dams across CONUS**. *Journal of Hydrology*:134145. **BibTeX** | **URL (https://www.sciencedirect.com/science/article/pii/S0022169425014830)**
EMNLP	Jayanth Krishna Chundru, Rudrashis Poddar, **Jie Cao**, and Tianyu Jiang.

2025. **Do LLMs Encode Frame Semantics? Evidence from Frame Identification**. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing(To Appear)*. Association for Computational Linguistics, November. **BibTeX** | **PDF (https://mlciv.com/papers/jayanth2025emnlp.pdf)**

| ICCV WORKSHOP | Songkun Yan, Zhi Li, Siyu Zhu, Yixin Wen, Mofan Zhang, Mengye Chen, **Jie Cao**, and Yang Hong. 2025. **AQUAH: Automatic Quantification and Unified Agent in Hydrology**. *1st workshop on Sustainability with Earth observation and AI (co-located with ICCV), 2025 (To Appear)*. **BibTeX** | **PDF (https://mlciv.com/papers/songkun2025sea.pdf)** | **URL (https://arxiv.org/abs/2508.02936)** |

| EDM | Jannatun Naim, **Jie Cao**, Fareen Tasneem, Jennifer Jacobs, Brent Milne, James Martin, and Tamara Sumner. 2025. **Towards Actionable Pedagogical Feedback: A Multi- Perspective Analysis of Mathematics Teaching and Tutoring Dialogue**. In *Proceedings of the 18th International Conference on Educational Data Mining*, pages 328–341. International Educational Data Mining Society, July. **BibTeX** | **PDF (https://mlciv.com/papers/naim2025edm.pdf)** |

| ARXIV | Yingheng Tang, Wenbin Xu, **Jie Cao**, Weilu Gao, Steve Farrell, Benjamin Erichson, Michael W Mahoney, Andy Nonaka, and Zhi Yao. 2025. **MatterChat: A Multi-Modal LLM for Material Science**. *arXiv preprint arXiv:2502.13107*. **BibTeX** | **PDF (https://mlciv.com/papers/tang2025matterchat.pdf)** |

| COLING | **Jie Cao**, Abhijit Suresh, Jennifer Jacobs, Charis Clevenger, Amanda Howard, Chelsea Brown, Brent Milne, Tom Fischaber, Tamara Sumner, and James H. Martin. 2025. **Enhancing Talk Moves Analysis in Mathematics Tutoring through Classroom Teaching Discourse**. In *The 31st International Conference on Computational Linguistics*. **BibTeX** | **PDF (https://mlciv.com/papers/talkmove-coling-2024.pdf)** |

| L@S | Baptiste Moreau-Pernet, Yu Tian, Sandra Sawaya, Peter Foltz, **Jie Cao**, Brent Milne, and Thomas Christie. 2024. **Classifying Tutor Discursive Moves at Scale in Mathematics Classrooms with Large Language Models**. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale*, pages 361–365. Association for Computing Machinery. **BibTeX** | **PDF (https://mlciv.com/papers/talkmove-llm-2024.pdf)** | **URL** |

(https://doi.org/10.1145/3657604.3664664)

IEEE TVCG   Zhimin Li, Shusen Liu, Xin Yu, Kailkhura Bhavya, **Jie Cao**, Diffenderfer James Daniel, Peer-Timo Bremer, and Valerio Pascucci. 2024. **"Understanding Robustness Lottery": A Geometric Visual Comparative Analysis of Neural Network Pruning Approaches**. *IEEE Transactions on Visualization and Computer Graphics.*   **BibTeX** | **PDF (https://mlciv.com/papers/vis-pruning2024.pdf)** | **URL (https://doi.org/10.1109/tvcg.2024.3514996)**

BEA   E. Margaret Perkoff, Abhidip Bhattacharyya, Jon Cai, and **Jie Cao**. 2023. **Comparing Neural Question Generation Architectures for Reading Comprehension**. In Ekaterina Kochmar, Jill Burstein, Andrea Horbach, Ronja Laarmann-Quante, Nitin Madnani, Anaïs Tack, Victoria Yaneva, Zheng Yuan, and Torsten Zesch, editors, *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 556–566, Toronto, Canada, July. Association for Computational Linguistics.   **BibTeX** | **PDF (https://mlciv.com/papers/qg-bea23.pdf)** | **URL (https://aclanthology.org/2023.bea-1.47/)**

UMAP   **Jie Cao**, Ananya Ganesh, Jon Cai, Rosy Southwell, Magerate Perkoff, Michael Regan, Katharina Kann, James Martin, Martha Palmer, and Sideny D'Mello. 2023. **A Comparative Analysis of Automatic Speech Recognition Errors in Small Group Classroom Discourse**. *Proceedings of the 31st ACM Conference on User Modeling Adaptation and Personalization.*   **BibTeX** | **PDF (https://mlciv.com/papers/cao-umap23.pdf)**

ACL   Ananya Ganesh, **Jie Cao**, E. Magerate Perkoff, Rosy Southwell, Martha Palmer, and Katharina Kann. 2023. **Mind the Gap between the Application Track and the Real World**. *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics, 2023.*   **BibTeX** | **PDF (https://mlciv.com/papers/ananya-acl23.pdf)**

IWSDS   Jon Cai, Brendan D. King, Margaret Perkoff, Shiran Dudy, **Jie Cao**, Marie Grace, Natalia Wojarnik, Ganesh Ananya, James Martin, Martha Palmer, Marilyn Walker, and Jeffrey Flanigan. 2022. **Dependency Dialogue Acts — Annotation Scheme and Case Study**. *The 13th International Workshop on Spoken Dialogue Systems Technology.*   **BibTeX** | **PDF (https://mlciv.com/papers/jon-dda2022.pdf)**

VLDB   Debjyoti Paul*, **Jie Cao***, Feifei Li, and Vivek Srikumar. 2021. **Database**

	Workload Characterization with Query Plan Encoders. *Proceedings of the VLDB Endowment*, 15(4):923–935. **BibTeX**	**PDF (https://mlciv.com/papers/cao2021dbqencoder.pdf)**
NAACL	**Jie Cao** and Yi Zhang. 2021. **A Comparative Study on Schema-Guided Dialogue State Tracking**. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 782–796. **BibTeX**	**PDF (https://mlciv.com/papers/cao2021comparative.pdf)**	**Poster (https://mlciv.com/papers/cao2021comparative.poster.pdf)**
ACL	Zhiqiang Liu, Zuohui Fu, **Jie Cao**, Gerard de Melo, Yik-Cheung Tam, Cheng Niu, and Jie Zhou. 2019. **Rhetorically Controlled Encoder-Decoder for Modern Chinese Poetry Generation**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. **BibTeX**	**PDF (https://mlciv.com/papers/rhetorical-poetry2019.pdf)**
CONLL	**Jie Cao**, Yi Zhang, Adel Youssef, and Vivek Srikumar. 2019. **Amazon at MRP 2019: Parsing Meaning Representations with Lexical and Phrasal Anchoring**. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the Conference on Natural Language Learning*, pages 138–148. **BibTeX**	**PDF (https://mlciv.com/papers/cao2019amazon.pdf)**
ACL	**Jie Cao**, Michael Tanana, Zac Imel, Eric Poitras, David Atkins, and Vivek Srikumar. 2019. **Observing Dialogue in Therapy: Categorizing and Forecasting Behavioral Codes**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. **BibTeX**	**PDF (https://mlciv.com/papers/cao2019observing.pdf)**	**Slides (https://mlciv.com/papers/cao2019observing.slides.pdf)**

Academic Service

- Standing Reviewer for Journals: **Computational Linguistics (https://submissions.cljournal.org/index.php/cljournal)**
- Area Chair for **ACL Rolling Review(ARR) (https://aclrollingreview.org/)**
- PC Member / Reviewer for Conferences and Workshops: ACL, EMNLP, NAACL, EACL, COLING, CoNLL, COLM, AAAI, ACL Rolling Review, AIED, EDM, **MRP'2019 (http://mrp.nlpl.eu/2019/)**, **BEA (https://aclanthology.org/venues/bea/)**, **NLP4ConvAI (https://www.aclweb.org/portal/content/call-papers-5th-workshop-nlp-conversational-ai)**, **AmericasNLP'23 (https://turing.iimas.unam.mx/americasnlp/2023_st.html)**, **SLaTE'23**

⌃

ABOUT THIS SITE

 (https://clustrmaps.com/site/1c1qk)

Jie Cao's Personal Website. **More › (https://mlciv.com/info/)**

SERVICES

()
Contact (/contact/)
AI-Deadlines (/ai-deadlines/)
RSS (/feed.xml)
Atom (/atom.xml)
sitemap.xml (/sitemap.xml)

RELATED LINKS

()
OU NLP (https://ounlp.org)
Utah NLP (https://nlp.cs.utah.edu)
iSAT (https://www.colorado.edu/research/ai-institute/)
CLEAR (https://www.colorado.edu/lab/clear/)