

## Education

---

2015 – 2022	<b>Doctor of Philosophy</b>
	Computer Science
	<i>University of Utah</i>
<i>Advisor:</i>	Vivek Srikumar
<i>Dissertation:</i>	Inductive Biases for Deep Linguistic Structured Prediction with Independent Factorization [6]

## Research Experience

---

NSF-iSAT, University of Colorado Boulder, Boulder 2022 - present

- I am a post-doctoral researcher at NSF AI Institute for Student-AI Teaming (iSAT). I worked closely with James Martin, Boulder NLP Groups, and Institute of Cognitive Science. Main research mainly covers conversational modelling [5, 4, 1] on various educational setting, including small group collaborative learning [3] and intelligent agents [2]

Utah NLP Group, University of Utah, Salt Lake City 2015 - 2022

- I worked with Prof. Vivek Srikumar in Utah NLP Group. I study methods for predicting structured representations of natural language text, e.g. Semantic Parsing [11], Dialogue System [13, 10, 8].

Intern at AWS AI, Amazon Lex, Remote Summer 2020

- Our paper on schema-guided dialog got accepted to NAACL 2021 [8]

Intern at AWS AI, Amazon Lex, Seattle Summer 2019

- Cross framework meaning representation parsing, ranked 1st on AMR on CoNLL 2019 shared task [11]

Intern at Tencent, Wechat AI, Palo Alto Summer 2018

- End-to-End Goal-oriented Dialogue System [13], ranked 2nd in DSTC7 track1
- Rhetorically Controlled Poetry Generation[12] got accepted to ACL 2019

CGCL Lab, HUST, Wuhan 2008 - 2012

- When I was a master student at HUST, I worked with Prof. Xia Xie and Prof. Hai Jin. My research are around large scale statistical computing, programming language [14, 15], system virtualization [17, 16]

## Teaching & Mentoring

---

2019-2020	Mentor	Bachelor Thesis (Tarun Sunkaraneni), U of Utah
Spring 2018	TA	Structured Prediction, U of Utah
Fall 2016, 2018	TA	Machine Learning, U of Utah
Fall 2009	Lecturer	Algorithm and Data Structure, HUST
2006-2008	Leader	Algorithm&Game Team, Unique Studio, HUST

## Work Experience

---

Assitant Researcher at Sohu RDC Lab, Beijing 2014 - 2015

- Hadoop, Spark, Data migration, Data security, Distributed machine learning.

Senior Software Engineer at Zun Club, Beijing 2013 - 2014

- Heterogeneous data intergration, Hotel recommendation system.

## Software Engineer at Baidu, Beijing

2012 - 2014

- Baidu Voice Assistant, Query analysis, Dialogue
- Mobile Search Anti-Attack Ecosystem, Speed optimization

## Intern at Alibaba, Hangzhou

2010 - 2011

- KV Storage, MySQL, Database Replication, Real-time Computing, Distributed Pub/Sub Data Pipeline.

## Honors & Awards

- 2019 CoNLL Shared Task, MRP, Top 1 on AMR
- 2019 DSTC7, Top 2 on track 1
- 2010 VMware Cloud Computing Innovation Cup, Top 50
- 2009 Google Android Innovative Idea Sharing Award
- 2007 “Computer World” Magazine Scholarship
- 2007 Microsoft Imagine Cup
  - Algorithm Challenge, Top 50 World-Wide
  - Visual Gaming Contest, Top 2 in China, 18th World Final
- 2006 HUST ACM Programming Contest, Top 3

## Publications

- [1] Jon Z Cai, Brendan King, Margaret Perkoff, Shiran Dudy, **Jie Cao**, Marie Grace, Natalia Wojarnik, Ananya Ganesh, James H Martin, Martha Palmer, et al. “Dependency Dialogue Acts–Annotation Scheme and Case Study”. In: *The 13th International Workshop on Spoken Dialogue Systems Technology* (2023).
- [2] **Jie Cao**, Rachel Dickler, Marie Grace, Alessandro Roncone, Leanne Hirshfield, Marilyn Walker, and Martha Palmer. “Designing an AI Partner for Jigsaw Classrooms”. In: *Workshop on Language-Based AI Character Interaction with Children* (2023).
- [3] **Jie Cao**, Ananya Ganesh, Jon Cai, Rosy Southwell, E Margaret Perkoff, Michael Regan, Katharina Kann, James H Martin, Martha Palmer, and Sidney D’Mello. “A Comparative Analysis of Automatic Speech Recognition Errors in Small Group Classroom Discourse”. In: *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 2023, pp. 250–262.
- [4] Ananya Ganesh, **Jie Cao**, E. Magerate Perkoff, Rosy Southwell, Martha Palmer, and Katharina Kann. “Mind the Gap between the Application Track and the Real World”. In: *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics, 2023* (2023).
- [5] E. Margaret Perkoff, Abhidip Bhattacharyya, Jon Cai, and **Jie Cao**. “Comparing Neural Question Generation Architectures for Reading Comprehension”. In: *18th Workshop on Innovative Use of NLP for Building Educational Applications, 2023* (2023).
- [6] Jie Cao. “Inductive Biases for Deep Linguistic Structured Prediction with Independent Factorization”. In: *Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global*. (2777357718) (2022).
- [7] Zhimin Li, Shusen Liu, Xin Yu, Kailkhura Bhavya, **Jie Cao**, Diffenderfer James Daniel, Peer-Timo Bremer, and Valerio Pascucci. “Understanding Robustness Lottery”: A Comparative Visual Analysis of Neural Network Pruning Approaches”. In: *arXiv preprint arXiv:2206.07918* (2022).
- [8] **Jie Cao** and Yi Zhang. “A Comparative Study on Schema-Guided Dialogue State Tracking”. In: *Proceedings of NAACL 2021*. 2021.
- [9] Debjyoti Paul\*, **Jie Cao\***, Feifei Li, and Vivek Srikumar. “Database Workload Characterization with Query Plan Encoders”. In: *VLDB’22* (2021).
- [10] **Jie Cao**, Michael Tanana, Zac Imel, Eric Poitras, David Atkins, and Vivek Srikumar. “Observing Dialogue in Therapy: Categorizing and Forecasting Behavioral Codes”. In: *Proceedings of ACL 2019*. Florence, Italy, 2019.

- [11] **Jie Cao**, Yi Zhang, Adel Youssef, and Vivek Srikumar. “Amazon at MRP 2019: Parsing Meaning Representations with Lexical and Phrasal Anchoring”. In: *Proceedings of the Shared Task on MRP at the CoNLL 2019*. 2019.
- [12] Zhiqiang Liu, Zuohui Fu, **Jie Cao**, Gerard de Melo, Yik-Cheung Tam, Cheng Niu, and Jie Zhou. “Rhetorically Controlled Encoder-Decoder for Modern Chinese Poetry Generation”. In: *Proceedings of ACL 2019*. Florence, Italy, 2019.
- [13] Shuo Sun\*, Yik-Cheung Tam\*, **Jie Cao\***, Can-Xiang Yan, Zuohui Fu, Cheng Niu, and Jie Zhou. “End-to-end Gated Self-attentive Memory Network for Dialog Response Selection”. In: *Proceedings of AAAI, DSTC7 Workshop*. 2019.
- [14] Xijiang Ke, Hai Jin, Xia Xie, and **Jie Cao**. “A distributed SVM method based on the iterative MapReduce”. In: *International Conference on Semantic Computing(ICSC)*. IEEE. 2015.
- [15] **Jie Cao\***, Xia Xie\*, Hai Jin, Xijiang Ke, and Wenzhi Cao. “JRBridge: A framework of large-scale statistical computing for R”. In: *Asia-Pacific on Services Computing Conference (APSCC)*. IEEE. 2012.
- [16] **Jie Cao**, Xia Xie, and Jin Hai. “A real-time scheduling system on embedded virtualization”. Patent ZL201110410689.1 (CN). 2011.
- [17] **Jie Cao**. “Cache Performance Evaluation on Xen Virtualization Platform”. In: *Undergraduate Thesis*. 2009.

## Talks

---

Slides for selected talks can be found in my slideshare site. <https://www.slideshare.net/jiessiecao/>

- 05/2021, Guest Lecturer @ University of Idaho: CS 501 Seminar: Contemporary Issues, ‘Inductive Biases in Deep Linguistic Structured Prediction’
- 06/2021, Conference Presentation on NAACL2021, ‘A Comparative Study on Schema-Guided Dialogue State Tracking’
- 12/2020, Talk on EMNLP’2020 Watch Party@Amazon Lex, ‘Task-oriented Conversational Semantic Parsing’
- 11/2019, Conference Presentation on MRP 2019, ‘Amazon at MRP 2019: Parsing Meaning Representations with Lexical and Phrasal Anchoring’
- 01/2016, Seminar Presentation on Spring’2016 NLP seminar at University of Utah, ‘Combinatory Categorical Grammar’
- 11/2015, Seminar Presentation on Fall’2015 Deep Learning seminar at University of Utah, ‘Parsing Natural Language Scense and Natural Language with Recursive Neural Networks’
- 11/2015, Seminar Presentation on Fall’2015 Data Science Seminar at University of Utah, ‘Challenges on Data Parallelism and Model Parallelism’
- 12/2015, Winner Presentation on Fall’2015 NLP Class at University of Utah, ‘Talking Geckos QA System’, our QA system ranked 1st in QA competition.

## Software

---

Other code for papers can be found in <https://github.com/utahnlp> and <https://github.com/mlciv>

- LAPA-MRP, our open-sourced system submission to MRP 2019 shared task at CoNLL 2019. Our model ranked 1st in the AMR subtask, 5th in UCCA, 6th in PSD and 7th in DM.
- dt-proxy, A nodejs-based token proxy to support distributed data copy between secure and insecure hadoop clusters.
- alfred-workflows, A home-brew repo for productive alfred workflows, including propbank query, display rotation, and so on.
- renjin, a JVM-based interpreter for the R language. <https://github.com/bedatadriven/renjin>

## **Academic Service**

---

### **Program Committee Member/Reviewer**

- 24th International Conference on Artificial Intelligence in Education (AIED), 2023
- Conference of the Association for the Advancement of Artificial Intelligence (AAAI), 2020–22;
- Annual meeting of the Association of Computational Linguistics (ACL), 2020–23;
- International Conference on Computational Linguistics (COLING), 2020;
- Conference of the European Chapter of the ACL (EACL), 2021;
- Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020–2022;
- North American Chapter of the ACL (NAACL), 2021;
- Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning (MRP’2019@CoNLL).
- ACL Rolling Review, 2021–23