KAI SUN

ks985@cornell.edu   (+1) (607) 319-9668   https://www.kaisun.org/

EDUCATION

Cornell University   *Aug. 2015 – Aug. 2021*
Ph.D., Computer Science
- Advisor: Claire Cardie
- Thesis: *Machine Reading Comprehension: Challenges and Approaches*

Shanghai Jiao Tong University   *Sept. 2011 – June 2015*
B.S., Computer Science (ACM Honors Class, Zhiyuan College)
- Ranked **1st** in the ACM Honors Class

INDUSTRY EXPERIENCE

Meta (formerly Facebook) Reality Labs, Remote Work   Sept. 2021 – Present
Research Scientist

- Working on conversational AI in the AR/VR world.

Facebook Conversational AI, Redmond, WA   June 2020 – Aug. 2020
Research Intern, with Shane Moon, Paul Crook, Stephen Roller, Bing Liu, Stephen Wang, and Honglei Liu

- Proposed Adding Chit-Chat to ENhance Task-ORiented dialogues (ACCENTOR).
- Proposed a flexible data construction approach, a dataset, and models for ACCENTOR.

Tencent AI Lab, Bellevue, WA   Jan. 2020 – May 2020
Research Intern, with Dian Yu, Jianshu Chen, and Dong Yu

- Proposed methods to extract multiple types of contextualized knowledge from scripts and inject them into a machine reader.

Tencent AI Lab, Bellevue, WA   May 2018 – Jan. 2019
Research Intern, with Dian Yu, Jianshu Chen, and Dong Yu

- Proposed a new machine reading comprehension (MRC) task with a dataset and models.
- Proposed three domain-independent strategies to improve non-extractive MRC, leading to new state-of-the-art results on 7 representative non-extractive MRC datasets from different domains.

Microsoft Research Asia, Beijing   Aug. 2014 – Feb. 2015
Research Intern, with Frank Soong and Lijuan Wang

- Built an RNN-based text-to-speech (TTS) synthesis system that supports multi-task learning.
- Investigated to control the age dimension of RNN-based TTS models by applying multi-task learning and matrix decomposition.
- Studied handling head stabilization and registration problem of 3D talking head by computational geometry and computer vision algorithms.

ACADEMIC RESEARCH EXPERIENCE

Cornell University   2016 – 2021
Research Assistant, with Claire Cardie

- Worked on improving the state-of-the-art and exploring new tasks for machine reading comprehension and dialogue understanding.

- Studied sentiment and belief detection in discussion forums and newswires.
- Lead Developer of the Cornell Chinese Belief and Sentiment Detection Systems
  - Ranked 1st in BeSt Evaluation at TAC 2016 and TAC 2017.
  - The system was employed in the TAC 2017 Cold Start KB task as part of the Tinkerbell team (Columbia+Cornell+JHU+RPI+Stanford+UIUC+Upenn) and ranked 1st in the Chinese sentiment slot filling task.

Shanghai Jiao Tong University 2013 – 2015
Research Assistant, with Kai Yu

- Worked on improving the state-of-the-art in tracking the state of spoken dialogue systems.
- Lead Developer of the SJTU Dialogue State Tracking Systems
  - Ranked 1st/2nd/3rd respectively in 3 tasks in DSTC-3.

INDEPENDENT RESEARCH EXPERIENCE

Self-Motivated Research on AI in Board Games 2006 – 2018

- Author of *Yixin*, a Gomoku and Renju AI
  - World champion AI, winner of the 13th, 14th, 15th, 16th, 17th, 18th, and 19th Gomocup.
  - Yixin was the first Gomoku and Renju AI that can compete at the human champion level. It beat Taiwan's Meijin title holder Lin Shu-Hsuan and the world Gomoku champion Rudolf Dupszki in 2017, and drew with world Renju champion Qi Guan in 2018.

PUBLICATIONS

Preprints

- **Kai Sun**. Digital Asset Valuation: A Study on Domain Names, Email Addresses, and NFTs. arXiv preprint. 2022.

Journal Articles

- **Kai Sun**, Dian Yu, Dong Yu, and Claire Cardie. Investigating Prior Knowledge for Challenging Chinese Machine Reading Comprehension. Transactions of the Association for Computational Linguistics (TACL). 2020. (Also an oral presentation at ACL 2020.)
- **Kai Sun**, Dian Yu, Jianshu Chen, Dong Yu, Yejin Choi, and Claire Cardie. DREAM: A Challenge Dataset and Models for Dialogue-Based Reading Comprehension. Transactions of the Association for Computational Linguistics (TACL). 2019. (Also an oral presentation at NAACL 2019.)
- **Kai Sun**, Qizhe Xie, and Kai Yu. Recurrent Polynomial Network for Dialogue State Tracking. Dialogue and Discourse (D&D). 2016.
- Kai Yu, Lu Chen, **Kai Sun**, Su Zhu, and Qizhe Xie. Evolvable Dialogue State Tracking for Statistical Dialogue Management. Frontiers of Computer Science. 2015.
- Kai Yu, **Kai Sun**, Lu Chen, and Su Zhu. Constrained Markov Bayesian Polynomial for Efficient Dialogue State Tracking. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP). 2015.
- Kai Yu, Lu Chen, Bo Chen, **Kai Sun**, and Su Zhu. Cognitive Technology in Task-Oriented Dialogue Systems – Concepts, Advances and Future. Chinese Journal of Computers. 2014. (*Invited paper*)

Peer-Reviewed Conference and Workshop Papers

- **Kai Sun***, Dian Yu*, Jianshu Chen, Dong Yu, and Claire Cardie. Improving Machine Reading Comprehension with Contextualized Commonsense Knowledge. Annual Meeting of the Association for Computational Linguistics (ACL). 2022. (*: equal contribution)

- Dian Yu, **Kai Sun**, Dong Yu, and Claire Cardie. Self-Teaching Machines to Read and Comprehend with Large-Scale Multi-Subject Question Answering Data. Findings of the Association for Computational Linguistics: EMNLP 2021 (EMNLP Findings). 2021.
- **Kai Sun**, Seungwhan Moon, Paul Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, and Claire Cardie. Adding Chit-Chat to Enhance Task-Oriented Dialogues. Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2021.
- Liang Xu, Hai Hu, Xuanwei Zhang, Lu Li, Chenjie Cao, Yudong Li, Yechen Xu, **Kai Sun**, Dian Yu, Cong Yu, Yin Tian, Qianqian Dong, Weitang Liu, Bo Shi, Yiming Cui, Junyi Li, Jun Zeng, Rongzhao Wang, Weijian Xie, Yanting Li, Yina Patterson, Zuoyu Tian, Yiwen Zhang, He Zhou, Shaoweihua Liu, Zhe Zhao, Qipeng Zhao, Cong Yue, Xinrui Zhang, Zhengliang Yang, Kyle Richardson, and Zhenzhong Lan. CLUE: A Chinese Language Understanding Evaluation Benchmark. The 28th International Conference on Computational Linguistics (COLING). 2020.
- Dian Yu*, **Kai Sun***, Claire Cardie, and Dong Yu. Dialogue-Based Relation Extraction. Annual Meeting of the Association for Computational Linguistics (ACL). 2020. (*: equal contribution)
- Xiaoman Pan*, **Kai Sun***, Dian Yu, Jianshu Chen, Heng Ji, Claire Cardie, and Dong Yu. Improving Question Answering with External Knowledge. EMNLP Workshop on Machine Reading for Question Answering (MRQA). 2019. (*: equal contribution)
- Hai Wang, Dian Yu, **Kai Sun**, Jianshu Chen, Dong Yu, David McAllester, and Dan Roth. Evidence Sentence Extraction for Machine Reading Comprehension. The SIGNLL Conference on Computational Natural Language Learning (CoNLL). 2019.
- Hai Wang, Dian Yu, **Kai Sun**, Jianshu Chen, and Dong Yu. Improving Pre-Trained Multilingual Model with Vocabulary Expansion. The SIGNLL Conference on Computational Natural Language Learning (CoNLL). 2019.
- **Kai Sun**, Dian Yu, Dong Yu, and Claire Cardie. Improving Machine Reading Comprehension with General Reading Strategies. Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). 2019.
- **Kai Sun**, Su Zhu, Lu Chen, Siqiu Yao, Xueyang Wu, and Kai Yu. Hybrid Dialogue State Tracking for Real World Human-to-Human Dialogues. Conference of the International Speech Communication Association (Interspeech). 2016.
- Qizhe Xie, **Kai Sun**, Su Zhu, Lu Chen, and Kai Yu. Recurrent Polynomial Network for Dialogue State Tracking with Mismatched Semantic Parsers. 16th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL). 2015.
- **Kai Sun**, Lu Chen, Su Zhu, and Kai Yu. A Generalized Rule Based Tracker for Dialogue State Tracking. IEEE Spoken Language Technology Workshop (SLT). 2014.
- Su Zhu, Lu Chen, **Kai Sun**, Da Zheng, and Kai Yu. Semantic Parser Enhancement for Dialogue Domain Extension with Little Data. IEEE Spoken Language Technology Workshop (SLT). 2014.
- **Kai Sun**, Lu Chen, Su Zhu, and Kai Yu. The SJTU System for Dialog State Tracking Challenge 2. 15th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL). 2014.

Non-Refereed Papers

- **Kai Sun** and Claire Cardie. Cornell Belief and Sentiment System at TAC 2017. Text Analysis Conference (TAC). 2017.
- Mohamed Al-Badrashiny, Jason Bolton, Arun Tejavsi Chaganty, Kevin Clark, Craig Harman, Lifu Huang, Matthew Lamm, Jinhao Lei, Di Lu, Xiaoman Pan, Ashwin Paranjape, Ellie Pavlick, Haoruo Peng, Peng Qi, Pushpendre Rastogi, Abigail See, **Kai Sun**, Max Thomas, Chen-Tse Tsai, Hao Wu, Bo-liang Zhang, Chris Callison-Burch, Claire Cardie, Heng Ji, Christopher Manning, Smaranda Muresan, Owen C. Rambow, Dan Roth, Mark Sammons, and Benjamin Van Durme. TinkerBell: Cross-lingual Cold-Start Knowledge Base Construction. Text Analysis Conference (TAC). 2017.

- Vlad Niculae, **Kai Sun**, Xilun Chen, Yao Cheng, Xinya Du, Esin Durmus, Arzoo Katiyar, and Claire Cardie. Cornell Belief and Sentiment System at TAC 2016. Text Analysis Conference (TAC). 2016.

INVITED TALKS

- *Dialogue-Based Relation Extraction.* Workshop on Personal Knowledge Graphs, AKBC'21, Oct. 2021.
- *The Cornpittmich Chinese System for BeSt Evaluation 2016.* 2016 Text Analysis Conference (TAC), NIST-Gaithersburg, MD, Nov. 2016.

HONORS AND AWARDS

- **University Fellowship**, Cornell University, 2015
- **Excellent Bachelor Thesis (Top 1%)**, Shanghai Jiao Tong University, 2015
- **Zhiyuan Excellent Student Scholarship**, Shanghai Jiao Tong University, 2015 (highest honor in Zhiyuan College)
- **Outstanding Graduate Award**, Shanghai Jiao Tong University, 2015
- **Award of Excellence of MSRA Star of Tomorrow Internship Program**, 2015
- **National Scholarship**, 2014 (highest scholarship in China, awarded to top 1% students)
- **Microsoft Young Fellow Scholarship Award**, 2014 (39 undergraduate&graduate students in China)
- **Google Excellence Scholarship**, 2014 (58 undergraduate&graduate students in China)
- **Shanghai Government Scholarship**, 2013 (1 student in the ACM Honors Class)
- **Chun-Tsung Scholarship**, 2013 (1 student in the ACM Honors Class)
- **Academic Excellence Scholarship (First-Class)**, Shanghai Jiao Tong University, 2012 (awarded to top 1% students)
- **KoGuan Scholarship**, 2012 (55 undergraduate students in SJTU)
- **First Prize**, National High School Math League, 2010
- **Silver Medal**, National Olympiad in Informatics (NOI), 2010

TEACHING EXPERIENCE

- Fall 2019: Teaching Assistant in CS 4740/5740 Introduction to Natural Language Processing
- Fall 2017: Teaching Assistant in CS 4740/5740 Introduction to Natural Language Processing
- Fall 2016: Teaching Assistant in CS 4740/5740 Introduction to Natural Language Processing
- Summer 2016: Teaching Assistant in CS 2110 Object-Oriented Programming and Data Structures
- Spring 2015: Teaching Fellow in MS 109 Compiler Design and Implementation (Lecturer)
- Fall 2013: Teaching Fellow in CS 101 Introduction to Computer Science (Lecturer)
- Fall 2012: Teaching Assistant in CS 101 Introduction to Computer Science (Lecturer)

PROFESSIONAL ACTIVITIES AND SERVICES

- Organizer: Gomocup (2016 – Present)
- Area Chair: AACL (2022)
- Program Committee Member: AAAI (2023), ACL (2019, 2020, 2021), COLING (2020, 2022), EACL (2023), EMNLP (2020, 2021, 2022), MRQA (2019, 2021), NAACL (2019, 2021), NLP4ConvAI (2021, 2022), NLPCC (2021, 2022), Spa-NLP (2022)
- Conference/Journal Reviewer: ACL ARR (2021, 2022), PRLETTERS (2022), TAI (2021), TASLP (2021), TOIS (2020)
- Book Proposal Reviewer: Elsevier (2020)