Lei Li

Contact Information

6403 Gates-Hillman Center, Carnegie Mellon University. Pittsburgh, PA 15213.
E-mail: leili@cs.cmu.edu
Homepage: `https://www.cs.cmu.edu/~leili`

Research Interests

Multilingual NLP and Machine Translation
Trustworthy Generative AI
Agentic Large Language Models
AI for Science: Drug Discovery, Protein Design

Education

Carnegie Mellon University, Pittsburgh, Pennsylvania USA

Ph.D., Computer Science Department, Sep. 2011
Dissertation: Fast algorithms for mining co-evolving time series.
Advisor: Prof. Christos Faloutsos

Shanghai Jiao Tong University, China

B.S., Computer Science (ACM Honored Class), Sep. 2002 - Jun. 2006
Thesis: Semantic Search based on Probabilistic Description Logic Program.
Advisor: Prof. Yong Yu

Employment

Carnegie Mellon University   **since Aug 2023**

Associate Professor, School of Computer Science, Language Technologies Institute, CyLab (affiliated), and CPCB (core faculty).

University of California, Santa Barbara   **since Jul. 2021**

Affiliated Faculty, Computer Science Department (Jul 2024 -)
Associate Professor (with tenure), Computer Science Department (Jul 2023 - Jun 2024)
Assistant Professor, Computer Science Department (Jul 2021 - Jun 2023)

ByteDance   **Mar. 2016 - Jul. 2021**

The Founding Director of ByteDance AI Lab
ByteDance is the most valuable startup (by Harvard Business Review in 2022) and the developer of Tiktok, Douyin, Toutiao, Xigua, and Lark Feishu with over 1 .5 billions of users in 150 countries.
Leading a group of over 120 researchers, engineers, and interns, on NLP, ML, Robotics, and Drug Discovery. Deployed systems: (1) VolcTrans Machine Translation system which serves over one billion users in 114 languages; (2) Key technologies for ads creation on OceanEngine, an online advertising system; (3) Query-analysis technologies for Toutiao Search, serving more than hundreds of millions of users daily; (4) Large-scale video infringement detection system with more than 20 millions of videos analyzed daily; (5) Zhuge, an E-Sales Assistant system; and (6) Xiaomingbot, an automatic news writing system.

Baidu USA   **Oct. 2014 - Mar. 2016**

Principal Research Scientist in Institute of Deep Learning, Baidu Research
focus on deep learning and natural language understanding, deployed AI components to Baidu ads system, and Baidu maps.

University of California, Berkeley   **Oct. 2011 - Oct. 2014**

Postdoctoral Researcher in the EECS department, working with Prof. Stuart Russell
Bayesian inference for open universe probability models. Developed BLOG interpreter v0.9 and Swift compiler for BLOG.

Carnegie Mellon University   **Oct. 2012 - Aug. 2013**

Visiting Researcher in the Machine Learning Department, working with Prof. Eric Xing

IBM T.J. Watson Research Center, Hawthorne, New York USA   **May. 2010 - Aug. 2010**

Intern at Healthcare Transformation group.
Mentor: Jimeng Sun

Microsoft Research, Redmond, Washington USA,   **Jun. 2009 - Aug. 2009**

Intern at Networked Embedded Computing group.
Mentor: Jie Liu and Suman Nath.

Google, Mountain View, California USA   **May. 2008 - Aug. 2008**

Intern at Adspam group.
Mentor: Tim Pan.

APEX Data & Knowledge Management Lab, Shanghai, China   **Jul. 2004 - Jun. 2006**

Research Assistant
Mentor: Jian Zhou and Lei Zhang

Microsoft Research Asia, Beijing, China   **Aug, 2005 - Oct, 2005**

Intern, on *semantic file system* project.
Mentor: Chenxi Lin

Awards and Highlights

1. ICLR 2023 Notable Area Chair.
2. ACL 2021 Best Paper Award (1 in 3350 submissions).
3. Winning 1st places in WMT20 Machine Translation Contest for 5 language directions: Chinese-English, German-English, French-German, English-Khmer, and English-Pashto, as well as 1st places in Parallel Corpus Filtering competition for Khmer and Pashto. November, 2020.
4. 2019 Young Elite Award by China Computer Federation. (6 winners in total)
5. 2017 2nd-class Wu Wen-tsün Award in AI for "Key technologies of Internet information summarization and machine writing", the premier award by China Association of Artificial Intelligence.
6. 2017 China Computer Federation Distinguished Speaker.
7. 2012 SIGKDD Doctoral Dissertation Award (runner-up), the Association for Computing Machinery (3 winners and runner-ups in total).
8. National Excellent Patent award in 2021, China, for patent 201811060837.X.
9. Led the development of VolcTrans, a machine translation system, with billions of requests daily, serving over a billion of active users.
10. Led the development of AI writing bot Xiaomingbot.
    - It automatically generates sports news about 17 leagues/series including the Rio Olympics (2016) and 2018 FIFA World Cup, with over 150,000 followers on social media. The work attracted over 100 coverage from news outlets including MIT Technology Review, Newsweek, ChinaDaily, and Y Combinator.

- It is also deployed in an online advertisement platform, which helps hundreds of thousands of small and mid-sized enterprises to create better advertisements.
- Xiaomingbot also automatically generates timely missing-person announcements in text and videos for the Toutiao alert program of finding missing people with crowdsourcing, leading to more than 10,000 missing people found during 2016-2019. Multiple news coverage including GlobalTimes and Xinhuanet.
- It automatically writes product promotion articles in the Poverty-Alleviation program to help farmers to boost agriculture product sales on Toutiao platform, helping 20,000 families. Multiple news coverage including Xinhuanet.

11. Led the development of a large scale video duplication and infringement detection system, serving over one billion active users and analyzing 20 million videos daily by the end of 2019.
12. Led the development of key technologies of query analysis for Toutiao Search engine, which serves hundreds of millions of users daily, bringing in significant revenue.

Funding

1. Hydrox AI gift fund. $150,000. Jan 2025.
2. NEC Faculty Research Award. $50,000. Feb 2025.
3. Amazon Grant on "LLM-RAG Evaluation" . $750,000. Co-PI (PI: Mona Diab). October 2024.
4. NEC Faculty Research Award. $25,000. Sep 2024.
5. Alpha Design AI Gift Fund. $100,000. Aug 2024.
6. NIH U01: "m6A-suite: an informatics pipeline and resource for elucidating roles of m6A epitranscriptome in cancer", $56,744. sub-award from University of Pittsburgh (PI: Yufei Huang). April 1, 2024 - March 31, 2025.
7. CMU CyLab Seed Fund: Security Attacks and Robust Defense against LLMs. Co-PI (with Steven Wu, Daphne Ippolito). $50,000 (my portion: $16,667). 2024/3 - 2025/2.
8. Apple Gift Fund. Total $30,000. Sep 2023.
9. Amazon Research Award. "Real-time robust simultaneous interpretation with few samples". PI. Total: $120,000 (including AWS credits). 7/1/2023 - 6/30/2024.
10. Software Impact Grant from the Institute for Energy Efficiency (IEE) and the Investment Group of Santa Barbara (IGSB): "Compact and Energy-efficient AI Software on Mobile Devices", PI (with Co-PI William Wang), Total: $100,000 (my portion: $50,000). 10/1/2022 - 6/30/2023.
11. UCSB Faculty Research Grant: "Efficient AI Methods for De-novo Drug Molecule Design", PI, $13,709, 7/1/2022 - 6/30/2023.
12. UCSB-IEE-Meta Collaborative Research on AI, co-PI, my portion: $151,380, 10/1/2021 - 6/30/2023.
13. NSF #2219627 Grant for AAAI Doctoral Consortium 2022, PI, $6,703.

Chapters in Books

[Ch1] **Lei Li**. "Deep Generative Models". In: *Artificial Intelligence for Drug Discovery*. Ed. by Honglin Li, Mingyue Zheng, Feng Zhu, and Fang Bai. in Chinese. Chemical Industry Press, 2023. Chap. 3, pp. 78–94.

Refereed Journal Papers – Published

[J7] Jielin Qiu, William Han, Xuandong Zhao, Shangbang Long, Christos Faloutsos, **Lei Li**. "Evaluating Durability: Benchmark Insights into Image and Text Watermarking". In: *Transactions on Machine Learning Research* (2025).

[J6] Xinlong Wang, Rufeng Zhang, Chunhua Shen, Tao Kong, **Lei Li**. "SOLO: A Simple Framework for Instance Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* 44.11 (Nov. 2022), pp. 8587–8601.

[J5] Ruihang Chu, Yukang Chen, Tao Kong, Lu Qi, **Lei Li**. "ICM-3D: Instantiated Category Modeling for 3D Instance Segmentation". In: *IEEE Robotics and Automation Letters (RA-L)* 7.1 (Jan. 2022), pp. 57–64.

[J4] Tao Kong, Fuchun Sun, Huaping Liu, Yuning Jiang, **Lei Li**, Jianbo Shi. "FoveaBox: Beyound Anchor-based Object Detection". In: *IEEE Transactions on Image Processing* 29 (June 2020), pp. 7389–7398.

[J3] Fei Wu, Cewu Lu, Mingjie Zhu, Hao Chen, Jun Zhu, Kai Yu, **Lei Li**, Ming Li, Qianfeng Chen, Xi Li, Xudong Cao, Zhongyuan Wang, Zhengjun Zha, Yueting Zhuang, Yunhe Pan. "Towards a new generation of artificial intelligence in China". In: *Nature Machine Intelligence* 2 (June 2020), pp. 312–316.

[J2] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, **Lei Li**, Christos Faloutsos. "Non-linear Dynamics of Information Diffusion in Social Networks". In: *ACM Transactions on the Web* 11.1 (Feb. 2017).

[J1] **Lei Li**, Bin Fu, Christos Faloutsos. "Efficient Parallel Learning of Hidden Markov chain Models on SMPs". In: *IEICE Transactions on Information and Systems* E93.D.6 (2010), pp. 1330–1342.

Refereed Conference Papers

For reference, the most prestigious conferences in each area and their acceptance rates are as follows. Natural language processing's top conferences are ACL, EMNLP, NAACL with acceptance rates ≈25%. Machine learning/AI are NeurIPS, ICML, ICLR, AAAI, IJCAI with acceptance rates ≈25%. Data mining is KDD with acceptance rates ≈15%. Computer vision are CVPR, ICCV, ECCV with acceptance rates ≈25%. System's top conference SC(SuperComputing) with acceptance rates ≈20%. Speech are ICASSP, InterSpeech with acceptance rates ≈50%. Robotics are ICRA, IROS with acceptance rates ≈50%.

[C169] Danqing Wang, Jianxin Ma, Fei Fang, **Lei Li**. "TypedThinker: Diversify Large Language Model Reasoning with Typed Thinking". In: *International Conference on Learning Representations (ICLR)*. Apr. 2025.

[C168] Wenda Xu, Rujun Han, Zifeng Wang, Long Le, Dhruv Madeka, **Lei Li**, William Yang Wang, Rishabh Agarwal, Chen-Yu Lee, Tomas Pfister. "Speculative Knowledge Distillation: Bridging the Teacher-Student Gap Through Interleaved Sampling". In: *International Conference on Learning Representations (ICLR)*. Apr. 2025.

[C167] Kexun Zhang, Weiran Yao, Zuxin Liu, Yihao Feng, Zhiwei Liu, Rithesh R N, Tian Lan, **Lei Li**, Renze Lou, Jiacheng Xu, Bo Pang, Yingbo Zhou, Shelby Heinecke, Silvio Savarese, Huan Wang, Caiming Xiong. "Diversity Empowers Intelligence: Integrating Expertise of Software Engineering Agents". In: *International Conference on Learning Representations (ICLR)*. Apr. 2025.

[C166] Xuandong Zhao, **Lei Li**, Yu-Xiang Wang. "Permute-and-Flip: An Optimally Robust and Watermarkable Decoder for LLMs". In: *International Conference on Learning Representations (ICLR)*. Apr. 2025.

[C165] Siqi Ouyang, Oleksii Hrinchuk, Zhehuai Chen, Vitaly Lavrukhin, Jagadeesh Balam, **Lei Li**, Boris Ginsburg. "Anticipating Future with Large Language Model for Simultaneous Machine Translation". In: *Proceedings of 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*. Apr. 2025.

[C164] Jian Xie, Kexun Zhang, Jiangjie Chen, Siyu Yuan, Kai Zhang, Yikai Zhang, **Lei Li**, Yanghua Xiao. "Revealing the Barriers of Language Agents in Planning". In: *Proceedings of 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*. Apr. 2025.

[C163] Fei Yuan, Chang Ma, Shuai Yuan, Qiushi Sun, **Lei Li**. "KS-Lottery: Finding Certified Lottery Tickets for Multilingual Transfer in Large Language Models". In: *Proceedings of 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*. Apr. 2025.

[C162] Kexun Zhang, Shang Zhou, Danqing Wang, William Yang Wang, **Lei Li**. "Scaling LLM Inference Efficiently with Optimized Sample Compute Allocation". In: *Proceedings of 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*. Apr. 2025.

[C161] Brian Tufts, Xuandong Zhao, **Lei Li**. "A Practical Examination of AI-Generated Text Detectors for Large Language Models". In: *Proceedings of 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL) - Findings*. Apr. 2025.

[C160] Xi Xu, Wenda Xu, Siqi Ouyang, **Lei Li**. "CA*: Addressing Evaluation Pitfalls in Computation-Aware Latency for Simultaneous Speech Translation". In: *Proceedings of 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL) - Findings*. Apr. 2025.

[C159] Kexun Zhang, Jane Dwivedi-Yu, Zhaojiang Lin, Yuning Mao, William Yang Wang, **Lei Li**, Yi-Chia Wang. "Extrapolating to Unknown Opinions Using LLMs". In: *Proceedings of the 31st International Conference on Computational Linguistics (COLING)*. Jan. 2025.

[C158] Zixian Huang, Wenhao Zhu, Gong Cheng, **Lei Li**, Fei Yuan. "MindMerger: Efficient Boosting LLM Reasoning in non-English Languages". In: *the 38th Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2024.

[C157] Xuandong Zhao, Kexun Zhang, Zihao Su, Saastha Vasan, Ilya Grishchenko, Christopher Kruegel, Giovanni Vigna, Yu-Xiang Wang, **Lei Li**. "Invisible Image Watermarks Are Provably Removable Using Generative AI". In: *the 38th Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2024.

[C156] Qingxiu Dong, **Lei Li**, Damai Dai, Ce Zheng, Jingyuan Ma, Rui Li, Heming Xia, Jingjing Xu, Zhiyong Wu, Baobao Chang, Xu Sun, **Lei Li**, Zhifang Sui. "A Survey on In-context Learning". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2024.

[C155] Danqing Wang, Kevin Yang, Hanlin Zhu, Xiaomeng Yang, Andrew Cohen, **Lei Li**, Yuandong Tian. "Learning Personalized Alignment for Evaluating Open-ended Text Generation". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2024.

[C154] Wenda Xu, Jiachen Li, William Yang Wang, **Lei Li**. "BPO: Supercharging Online Preference Learning by Adhering to the Proximity of Behavior LLM". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2024.

[C153] André Duarte, João Marques, Miguel Graça, Miguel Freire, **Lei Li**, Arlindo Oliveira. "LumberChunker: Long-Form Narrative Document Segmentation". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2024.

[C152] Yinquan Lu, Wenhao Zhu, **Lei Li**, Yu Qiao, Fei Yuan. "LLaMAX: Scaling Linguistic Horizons of LLM by Enhancing Translation Capabilities Beyond 100 Languages". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2024.

[C151] Jielin Qiu, Andrea Madotto, Zhaojiang Lin, Paul A. Crook, Yifan Ethan Xu, Xin Luna Dong, Christos Faloutsos, **Lei Li**, Babak Damavandi, Seungwhan Moon. "SnapNTell: Enhancing Entity-Centric Visual Question Answering with Retrieval Augmented Multimodal LLM". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2024.

[C150] Danqing Wang, Antonis Antoniades, Kha-Dinh Luong, Edwin Zhang, Mert Kosan, Jiachen Li, Ambuj Singh, William Yang Wang, **Lei Li**. "Global Human-guided Counterfactual Explanations for Molecular Properties via Reinforcement Learning". In: *the 30th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. Aug. 2024.

[C149] Fei Yuan, Shuai Yuan, Zhiyong Wu, **Lei Li**. "How Vocabulary Sharing Facilitates Multilingualism in LLaMA?" In: *the 62nd Annual Meeting of the Association for Computational Linguistics - Findings (ACL-Findings)*. Aug. 2024.

[C148] Kexun Zhang, Yee Man Choi, Zhenqiao Song, Taiqi He, William Yang Wang, **Lei Li**. "Hire a Linguist!: Learning Endangered Languages in LLMs with In-Context Linguistic Descriptions". In: *the 62nd Annual Meeting of the Association for Computational Linguistics - Findings (ACL-Findings)*. Aug. 2024.

[C147] Wenda Xu, Guanglei Zhu, Xuandong Zhao, Liangming Pan, **Lei Li**, William Yang Wang. "Pride and Prejudice: LLM Amplifies Self-Bias in Self-Refinement". In: *the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*. 2024. Aug. 2024.

[C146] André V. Duarte, Xuandong Zhao, Arlindo L. Oliveira, **Lei Li**. "DE-COP: Detecting Copyrighted Content in Language Models Training Data". In: *Proceedings of the 41st International Conference on Machine Learning (ICML)*. July 2024.

[C145] Zhenqiao Song, Yunlong Zhao, Wenxian Shi, Wengong Jin, Yang Yang, **Lei Li**. "Generative Enzyme Design Guided by Functionally Important Sites and Small-Molecule Substrates". In: *Proceedings of the 41st International Conference on Machine Learning (ICML)*. July 2024.

[C144] Zhenqiao Song, Tinglin Huang, **Lei Li**, Wengong Jin. "SurfPro: Functional Protein Design Based on Continuous Surface". In: *Proceedings of the 41st International Conference on Machine Learning (ICML)*. July 2024.

[C143] Wenda Xu, Daniel Deutsch, Mara Finkelstein, Juraj Juraska, Biao Zhang, Zhongtao Liu, William Yang Wang, **Lei Li**, Markus Freitag. "LLMRefine: Pinpointing and Refining Large Language Models via Fine-Grained Actionable Feedback". In: *Proceedings of 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) - Findings*. June 2024.

[C142] Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Lingpeng Kong, Jiajun Chen, **Lei Li**, Shujian Huang. "Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis". In: *Proceedings of 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) - Findings*. June 2024.

[C141] Xuandong Zhao, Prabhanjan Vijendra Ananth, **Lei Li**, Yu-Xiang Wang. "Provable Robust Watermarking for AI-Generated Text". In: *International Conference on Learning Representations (ICLR)*. May 2024.

[C140] Sameer Jain, Sedrick Scott Keh, Shova Chhetri, Karun Dewan, Pablo Izquierdo, Johanna Prussmann, Pooja Shrestha, César Suárez, Zheyuan Ryan Shi, **Lei Li**, Fei Fang. "Where It Really Matters: Few-Shot Environmental Conservation Media Monitoring for Low-Resource Languages". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2024.

[C139] Qingxiu Dong, Jingjing Xu, Lingpeng Kong, Zhifang Sui, **Lei Li**. "Statistical Knowledge Assessment for Large Language Models". In: *the 37th Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2023.

[C138] Kexun Zhang, Danqing Wang, Jingtao Xia, William Yang Wang, **Lei Li**. "ALGO: Synthesizing Algorithmic Programs with Generated Oracle Verifiers". In: *the 37th Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2023.

[C137] Danqing Wang, **Lei Li**. "Learning from Mistakes via Cooperative Study Assistant for Large Language Models". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Dec. 2023.

[C136] Wenda Xu, Danqing Wang, Liangming Pan, Zhenqiao Song, Markus Freitag, William Yang Wang, **Lei Li**. "INSTRUCTSCORE: Explainable Text Generation Evaluation with Finegrained Feedback". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Dec. 2023.

[C135] Siqi Ouyang, **Lei Li**. "AutoPlan: Automatic Planning of Interactive Decision-Making Tasks With Large Language Models". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Dec. 2023.

[C134] Bohong Wu, Fei Yuan, Hai Zhao, **Lei Li**, Jingjing Xu. "Extrapolating Multilingual Understanding Models as Multilingual Generators". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Dec. 2023.

[C133] Danqing Wang, Zeyu Wen, Fei Ye, **Lei Li**, Hao Zhou. "Accelerating Antimicrobial Peptide Discovery with Latent Structure". In: *the 29th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. Aug. 2023.

[C132] Zhenqiao Song, **Lei Li**. "Importance Weighted Expectation-Maximization for Protein Sequence Design". In: *Proceedings of the 40th International Conference on Machine Learning (ICML)*. July 2023.

[C131] Kexun Zhang, Xianjun Yang, William Yang Wang, **Lei Li**. "ReDi: Efficient Learning-Free Diffusion Inference via Trajectory Retrieval". In: *Proceedings of the 40th International Conference on Machine Learning (ICML)*. July 2023.

[C130] Xuandong Zhao, Yu-Xiang Wang, **Lei Li**. "Protecting Language Generation Models via Invisible Watermarking". In: *Proceedings of the 40th International Conference on Machine Learning (ICML)*. July 2023.

[C129] Fei Yuan, Yinquan Lu, Wenhao Zhu, Lingpeng Kong, **Lei Li**, Yu Qiao, Jingjing Xu. "Lego-MT: Learning Detachable Models for Massively Multilingual Machine Translation". In: *the 61st Annual Meeting of the Association for Computational Linguistics - Findings (ACL-Findings)*. July 2023.

[C128] Jiangjie Chen, Wei Shi, Ziquan Fu, Sijie Cheng, **Lei Li**, Yanghua Xiao. "Say What You Mean! Large Language Models Speak Too Positively about Negative Commonsense Knowledge". In: *the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*. July 2023.

[C127] Siqi Ouyang, Rong Ye, **Lei Li**. "WACO: Word-Aligned Contrastive Learning for Speech Translation". In: *the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*. July 2023.

[C126] Wenda Xu, Xian Qian, Mingxuan Wang, **Lei Li**, William Yang Wang. "SESCORE2: Learning Text Generation Evaluation via Synthesizing Realistic Mistakes". In: *the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*. July 2023.

[C125] Xuandong Zhao, Siqi Ouyang, Zhiguo Yu, Ming Wu, **Lei Li**. "Pre-trained Language Models can be Fully Zero-Shot Learners". In: *the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*. July 2023.

[C124] Jiangjie Chen, Rui Xu, Wenxuan Zeng, Changzhi Sun, **Lei Li**, Yanghua Xiao. "Converge to the Truth: Factual Error Correction via Iterative Constrained Editing". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2023.

[C123] Qingxiu Dong, Damai Dai, Yifan Song, Jingjing Xu, Zhifang Sui, **Lei Li**. "Calibrating Factual Knowledge in Pretrained Language Models". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Dec. 2022.

[C122] Wenda Xu, Yilin Tuan, Yujie Lu, Michael Saxon, **Lei Li**, William Yang Wang. "Not All Errors are Equal: Learning Text Generation Metrics using Stratified Error Synthesis". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Dec. 2022.

[C121] Xuandong Zhao, **Lei Li**, Yu-Xiang Wang. "Distillation-Resistant Watermarking for Model Protection in NLP". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Dec. 2022.

[C120] Xiaohui Wang, Yang Wei, Ying Xiong, Guyue Huang, Xian Qian, Yufei Ding, Mingxuan Wang, **Lei Li**. "LightSeq2: Accelerated Training for Transformer-based Models on GPUs". In: *Proceedings of The International Conference for High Performance Computing, Networking, Storage and Analysis (SC'22)*. Nov. 2022.

[C119] Yunfei Lu, Peng Cui, Linyun Yu, **Lei Li**, Wenwu Zhu. "Uncovering the Heterogeneous Effects of Preference Diversity on User Activeness: A Dynamic Mixture Model". In: *the 28th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. Aug. 2022.

[C118] Yiran Chen, Zhenqiao Song, Xianze Wu, Danqing Wang, Jingjing Xu, Jiaze Chen, Hao Zhou, **Lei Li**. "MTG: A Benchmark Suite for Multilingual Text Generation". In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT Findings)*. Association for Computational Linguistics, July 2022.

[C117] Rong Ye, Mingxuan Wang, **Lei Li**. "Cross-modal Contrastive Learning for Speech Translation". In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. Association for Computational Linguistics, July 2022.

[C116] Xuandong Zhao, **Lei Li**, Yu-Xiang Wang. "Provably Confidential Language Modelling". In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. Association for Computational Linguistics, July 2022.

[C115] Fei Huang, Tianhua Tao, Hao Zhou, **Lei Li**, Minlie Huang. "On the Learning of Non-autoregressive Transformers". In: *Proceedings of the 39th International Conference on Machine Learning (ICML)*. July 2022.

[C114] Yunfei Li, Tao Kong, **Lei Li**, Yi Wu. "Learning Design and Construction with Varying-Sized Materials via Prioritized Memory Resets". In: *IEEE International Conference on Robotics and Automation (ICRA)*. May 2022.

[C113] Yu Bao, Hao Zhou, Shujian Huang, Dongqi Wang, Lihua Qian, Xinyu Dai, Jiajun Chen, **Lei Li**. "latent-GLAT: Glancing at Latent Variables for Parallel Text Generation". In: *the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*. May 2022.

[C112] Qianqian Dong, Yaoming Zhu, Mingxuan Wang, **Lei Li**. "Learning When to Translate for Streaming Speech". In: *the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*. May 2022.

[C111] Qingkai Fang, Rong Ye, **Lei Li**, Yang Feng, Mingxuan Wang. "STEMM: Self-learning with Speech-text Manifold Mixup for Speech Translation". In: *the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*. May 2022.

[C110] Zhiyi Fu, Wangchunshu Zhou, Jingjing Xu, Hao Zhou, **Lei Li**. "Contextual Representation Learning beyond Masked Language Modeling". In: *the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*. May 2022.

[C109] Jiangjie Chen, Rui Xu, Ziquan Fu, Wei Shi, Zhongqiao Li, Xinbo Zhang, Changzhi Sun, **Lei Li**, Yanghua Xiao, Hao Zhou. "E-KAR: A Benchmark for Rationalizing Natural Language Analogical Reasoning". In: *the 60th Annual Meeting of the Association for Computational Linguistics (ACL) - Findings*. May 2022.

[C108] Zewei Sun, Mingxuan Wang, Hao Zhou, Chengqi Zhao, Shujian Huang, Jiajun Chen, **Lei Li**. "Rethinking Document-level Neural Machine Translation". In: *the 60th Annual Meeting of the Association for Computational Linguistics (ACL) - Findings*. May 2022.

[C107] Xuandong Zhao, Zhiguo Yu, Ming Wu, **Lei Li**. "Compressing Sentence Representation via Homomorphic Projective Distillation". In: *the 60th Annual Meeting of the Association for Computational Linguistics (ACL) - Findings*. May 2022.

[C106] Zhenqiao Song, Hao Zhou, Lihua Qian, Jingjing Xu, Shanbo Cheng, Mingxuan Wang, **Lei Li**. "switch-GLAT: Multilingual Parallel Machine Translation via Code-switch Decoder". In: *International Conference on Learning Representations (ICLR)*. Apr. 2022.

[C105] Huiyun Yang, Huadong Chen, Hao Zhou, **Lei Li**. "Enhancing Cross-lingual Transfer by Manifold Mixup". In: *International Conference on Learning Representations (ICLR)*. Apr. 2022.

[C104] Jiangjie Chen, Qiaoben Bao, Changzhi Sun, Xinbo Zhang, Jiaze Chen, Hao Zhou, Yanghua Xiao, **Lei Li**. "LOREN: Logic-Regularized Reasoning for Interpretable Fact Verification". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2022.

[C103] Jiangjie Chen, Chun Gan, Sijie Cheng, Hao Zhou, Yanghua Xiao, **Lei Li**. "Unsupervised Editing for Counterfactual Stories". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2022.

[C102] Chenyang Huang, Hao Zhou, Osmar Zaiane, Lili Mou, **Lei Li**. "Non-Autoregressive Translation with Layer-Wise Prediction and Deep Supervision". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2022.

[C101] Zaixiang Zheng, Hao Zhou, Shujian Huang, Jiajun Chen, Jingjing Xu, **Lei Li**. "Duplex Sequence-to-Sequence Learning for Reversible Machine Translation". In: *the 35th Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2021.

[C100] Qingnan Jiang, Mingxuan Wang, Jun Cao, Shanbo Cheng, Shujian Huang, **Lei Li**. "Learning Kernel-Smoothed Machine Translation with Retrieved Examples". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2021.

[C99] Dongyu Ru, Changzhi Sun, Jiangtao Feng, Lin Qiu, Hao Zhou, Weinan Zhang, Yong Yu, **Lei Li**. "Learning Logic Rules for Document-level Relation Extraction". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2021.

[C98] Zhiyuan Zeng, Jiaze Chen, Weiran Xu, **Lei Li**. "Gradient-based Adversarial Factual Consistency Evaluation for Abstractive Summarization". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2021.

[C97] Zewei Sun, Mingxuan Wang, **Lei Li**. "Multilingual Translation via Grafting Pre-trained Language Models". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2021.

[C96] Tao Wang, Chengqi Zhao, Mingxuan Wang, **Lei Li**, Hang Li, Deyi Xiong. "Secoco: Self-Correcting Encoding for Neural Machine Translation". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2021.

[C95] Yaoming Zhu, Jiangtao Feng, Chengqi Zhao, Mingxuan Wang, **Lei Li**. "Counter-Interference Adapter for Multilingual Machine Translation". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2021.

[C94] Danqing Wang, Jiaze Chen, Xianze Wu, Hao Zhou, **Lei Li**. "CNewSum: A Large-scale Chinese News Summarization Dataset with Human-annotated Adequacy and Deducibility Level". In: *The 10th CCF International Conference on Natural Language Processing and Chinese Computing (NLPCC)*. Qingdao, China, Oct. 2021.

[C93] Yunfei Li, Tao Kong, **Lei Li**, Yifeng Li, Yi Wu. "Learning to Design and Construct Bridge without Blueprint". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Sept. 2021.

[C92] Yiming Li, Tao Kong, Ruihang Chu, Yifeng Li, Peng Wang, **Lei Li**. "Simultaneous Semantic and Collision Learning for 6-DoF Grasp Pose Estimation". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Sept. 2021.

[C91] Wenxian Shi, Yuxuan Song, Hao Zhou, Bohan Li, **Lei Li**. "Follow Your Path: a Progressive Method for Knowledge Distillation". In: *Proc. of ECML-PKDD*. Sept. 2021.

[C90] Rong Ye, Mingxuan Wang, **Lei Li**. "End-to-end Speech Translation via Cross-modal Progressive Training". In: *Proc. of INTERSPEECH*. Aug. 2021.

[C89] Zehui Lin, Liwei Wu, Mingxuan Wang, **Lei Li**. "Learning Language Specific Sub-network for Multilingual Machine Translation". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*. Aug. 2021.

[C88] Xiao Pan, Liwei Wu, Mingxuan Wang, **Lei Li**. "Contrastive Learning for Many-to-many Multilingual Neural Machine Translation". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*. Aug. 2021.

[C87] Lihua Qian, Hao Zhou, Yu Bao, Mingxuan Wang, Lin Qiu, Weinan Zhang, Yong Yu, **Lei Li**. "Glancing Transformer for Non-Autoregressive Neural Machine Translation". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*. Aug. 2021.

[C86] Yijun Wang, Changzhi Sun, Yuanbin Wu, Hao Zhou, **Lei Li**, Junchi Yan. "UniRE: A Unified Label Space for Entity Relation Extraction". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*. Aug. 2021.

[C85] Runxin Xu, Tianyu Liu, **Lei Li**, Baobao Chang. "Document-level Event Extraction via Heterogeneous Graph-based Interaction Model with a Tracker". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*. Aug. 2021.

[C84] Jingjing Xu, Hao Zhou, Chun Gan, Zaixiang Zheng, **Lei Li**. "Vocabulary Learning via Optimal Transport for Neural Machine Translation". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*. Aug. 2021. ACL Best Paper Award (1/3350).

[C83] Chi Han, Mingxuan Wang, Heng Ji, **Lei Li**. "Learning Shared Semantic Space for Speech-to-Text Translation". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL) - Findings*. Aug. 2021.

[C82] Changzhi Sun, Xinbo Zhang, Jiangjie Chen, Chun Gan, Yuanbin Wu, Jiaze Chen, Hao Zhou, **Lei Li**. "Probabilistic Graph Reasoning for Natural Proof Generation". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL) - Findings*. Aug. 2021.

[C81] Danqing Wang, Jiaze Chen, Hao Zhou, Xipeng Qiu, **Lei Li**. "Contrastive Aligned Joint Learning for Multilingual Summarization". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL) - Findings*. Aug. 2021.

[C80] Liwei Wu, Shanbo Cheng, Mingxuan Wang, **Lei Li**. "Language Tags Matter for Zero-Shot Neural Machine Translation". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL) - Findings*. Aug. 2021.

[C79] Mingxuan Jing, Wenbing Huang, Fuchun Sun, Xiaojian Ma, Tao Kong, Chuang Gan, **Lei Li**. "Adversarial Option-Aware Hierarchical Imitation Learning". In: *Proceedings of the 38th International Conference on Machine Learning (ICML)*. July 2021.

[C78] Quanyu Long, Mingxuan Wang, **Lei Li**. "Generative Imagination Elevates Machine Translation". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. Online: Association for Computational Linguistics, June 2021, pp. 5738–5748.

[C77] Tao Wang, Chengqi Zhao, Mingxuan Wang, **Lei Li**, Deyi Xiong. "Autocorrect in the Process of Translation — Multi-task Learning Improves Dialogue Machine Translation". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers (NAACL-HLT)*. Online: Association for Computational Linguistics, June 2021, pp. 105–112.

[C76] Mingxuan Wang, Hongxiao Bai, Hai Zhao, **Lei Li**. "Cross-lingual Supervision Improves Unsupervised Neural Machine Translation". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers (NAACL-HLT)*. Online: Association for Computational Linguistics, June 2021, pp. 89–96.

[C75] Xiaohui Wang, Ying Xiong, Yang Wei, Mingxuan Wang, **Lei Li**. "LightSeq: A High Performance Inference Library for Transformers". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers (NAACL-HLT)*. Online: Association for Computational Linguistics, June 2021, pp. 113–120.

[C74] Yukang Chen, Yanwei Li, Tao Kong, Lu Qi, Ruihang Chu, **Lei Li**, Jiaya Jia. "Scale-aware Automatic Augmentation for Object Detection". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021.

[C73] Ya Jing, Tao Kong, Wei Wang, Liang Wang, **Lei Li**, Tieniu Tan. "Locate then Segment: A Strong Pipeline for Referring Image Segmentation". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021.

[C72] Peize Sun, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka, **Lei Li**, Zehuan Yuan, Changhu Wang, Ping Luo. "Sparse R-CNN: End-to-End Object Detection with Learnable Proposals". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021.

[C71] Xinlong Wang, Rufeng Zhang, Chunhua Shen, Tao Kong, **Lei Li**. "Dense Contrastive Learning for Self-Supervised Visual Pre-Training". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021. Oral.

[C70] Yutong Xie, Chence Shi, Hao Zhou, Yuwei Yang, Weinan Zhang, Yong Yu, **Lei Li**. "MARS: Markov Molecular Sampling for Multi-objective Drug Discovery". In: *International Conference on Learning Representations (ICLR)*. May 2021. Spotlight, 5.6% acceptance rate.

[C69] Yijun Wang, Changzhi Sun, Yuanbin Wu, Hao Zhou, **Lei Li**, Junchi Yan. "ENPAR: Enhancing Entity and Entity Pair Representations for Joint Entity Relation Extraction". In: *Proceedings of European Chapter of the Association for Computational Linguistics (EACL)*. Apr. 2021.

[C68] Zhenqiao Song, Jiaze Chen, Hao Zhou, **Lei Li**. "Triangular Bidword Generation for Sponsored Search Auction". In: *Proceedings of the 14th International Conference on Web Search and Data Mining (WSDM)*. Mar. 2021.

[C67] Qianqian Dong, Mingxuan Wang, Hao Zhou, Shuang Xu, Bo Xu, **Lei Li**. "Consecutive Decoding for Speech-to-text Translation". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2021.

[C66] Qianqian Dong, Rong Ye, Mingxuan Wang, Hao Zhou, Shuang Xu, Bo Xu, **Lei Li**. "Listen, Understand and Translate: Triple Supervision Decouples End-to-end Speech-to-text Translation". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2021.

[C65] Xunpeng Huang, Runxin Xu, Hao Zhou, Zhe Wang, Zhengyang Liu, **Lei Li**. "ACMo: Angle-Calibrated Moment Methods for Stochastic Optimization". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2021.

[C64] Jianze Liang, Chengqi Zhao, Mingxuan Wang, Xipeng Qiu, **Lei Li**. "Finding Sparse Structure for Domain Specific Neural Machine Translation". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2021.

[C63] Qingyang Wu, **Lei Li**, Zhou Yu. "TextGAIL: Generative Adversarial Imitation Learning for Text Generation". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2021.

[C62] Jieyu Zhang, Xiangchen Song, Ying Zeng, Jiaze Chen, Jiaming Shen, Yuning Mao, **Lei Li**. "Taxonomy Completion via Triplet Matching Network". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2021.

[C61] Xinlong Wang, Rufeng Zhang, Tao Kong, **Lei Li**, Chunhua Shen. "SOLOv2: Dynamic and Fast Instance Segmentation". In: *the 34th Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2020.

[C60] Bohan Li, Hao Zhou, Junxian He, Mingxuan Wang, Yiming Yang, **Lei Li**. "On the Sentence Embeddings from Pre-trained Language Models". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2020.

[C59] Zehui Lin, Xiao Pan, Mingxuan Wang, Xipeng Qiu, Jiangtao Feng, Hao Zhou, **Lei Li**. "Pre-training Multilingual Neural Machine Translation by Leveraging Alignment Information". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2020.

[C58] Shuang Zeng, Runxin Xu, Baobao Chang, **Lei Li**. "Double Graph Based Reasoning for Document-level Relation Extraction". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2020.

[C57] Dongyu Ru, Jiangtao Feng, Lin Qiu, Hao Zhou, Mingxuan Wang, Weinan Zhang, Yong Yu, **Lei Li**. "Active Sentence Learning by Adversarial Uncertainty Sampling in Discrete Space". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2020.

[C56] Maosen Zhang, Nan Jiang, **Lei Li**, Yexiang Xue. "Language Generation via Combinatorial Constraint Satisfaction: A Tree Search Enhanced Monte-Carlo Approach". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings*. Nov. 2020.

[C55] Yuxuan Song, Ning Miao, Hao Zhou, Lantao Yu, Mingxuan Wang, **Lei Li**. "Improving Maximum Likelihood Training for Text Generation with Density Ratio Estimation". In: *The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*. Aug. 2020.

[C54] Xinlong Wang, Tao Kong, Chunhua Shen, Yuning Jiang, **Lei Li**. "SOLO: Segmenting Objects by Locations". In: *The European Conference on Computer Vision (ECCV)*. Aug. 2020.

[C53] Wenxian Shi, Hao Zhou, Ning Miao, **Lei Li**. "Dispersed Exponential Family Mixture VAEs for Interpretable Text Generation". In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*. July 2020.

[C52] Ning Miao, Yuxuan Song, Hao Zhou, **Lei Li**. "Do you have the right scissors? Tailoring Pre-trained Language Models via Monte-Carlo Methods". In: *the 58th Annual Meeting of the Association for Computational Linguistics (ACL) - short papers*. July 2020.

[C51] Xinyu Hua, **Lei Li**, Lifeng Hua, Lu Wang. "XREF: Entity Linking for Chinese News Comments with Supplementary Article Reference". In: *Automated Knowledge Base Construction (AKBC)*. June 2020.

[C50] Rong Ye, Wenxian Shi, Hao Zhou, Zhongyu Wei, **Lei Li**. "Variational Template Machine for Data-to-Text Generation". In: *International Conference on Learning Representations (ICLR)*. Apr. 2020.

[C49] Zaixiang Zheng, Hao Zhou, Shujian Huang, **Lei Li**, Xinyu Dai, Jiajun Chen. "Mirror Generative Models for Neural Machine Translation". In: *International Conference on Learning Representations (ICLR)*. Apr. 2020. Oral, 1.9% acceptance rate.

[C48] Xunpeng Huang, Xianfeng Liang, Zhengyang Liu, **Lei Li**, Yue Yu, Yitan Li. "SPAN: A Stochastic Projected Approximate Newton Method". In: *the 34th AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2020.

[C47] Xinlong Wang, Wei Yin, Tao Kong, Yuning Jiang, **Lei Li**, Chunhua Shen. "Task-Aware Monocular Depth Estimation for 3D Object Detection". In: *the 34th AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2020. Oral.

[C46] Qingyang Wu, **Lei Li**, Hao Zhou, Ying Zeng, Zhou Yu. "Importance-Aware Learning for Neural Headline Editing". In: *the 34th AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2020.

[C45] Jiacheng Yang, Mingxuan Wang, Hao Zhou, Chengqi Zhao, Weinan Zhang, Yong Yu, **Lei Li**. "Towards Making the Most of BERT in Neural Machine Translation". In: *the 34th AAAI Conference on Artificial Intelligence (AAAI)*. Feb. 2020.

[C44] Ning Miao, Hao Zhou, Chengqi Zhao, Wenxian Shi, **Lei Li**. "Kernelized Bayesian Softmax for Text Generation". In: *the 33rd Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2019.

[C43] Zhichen Zhao, **Lei Li**, Bowen Zhang, Meng Wang, Yuning Jiang, Li Xu, Fengkun Wang, Weiying Ma. "What You Look Matters: Offline Evaluation of Advertising Creatives for Cold Start Problem". In: *the 28th ACM International Conference on Information and Knowledge Management (CIKM)*. Nov. 2019.

[C42] Mingxuan Wang, Jun Xie, Zhixing Tan, Jinsong Su, Deyi Xiong, **Lei Li**. "Towards Linear Time Neural Machine Translation with Capsule Networks". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2019.

[C41] Qing-Yuan Jiang, Yi He, Gen Li, Jian Lin, **Lei Li**, Wu-Jun Li. "SVD: A Large-Scale Short Video Dataset for Near Duplicate Video Retrieval." In: *International Conference on Computer Vision (ICCV)*. Oct. 2019.

[C40] Xin Wang, Jiawei Wu, Junkun Chen, **Lei Li**, Yuan-Fang Wang, William Yang Wang. "VATEX: A Large-Scale, High-Quality Multilingual Dataset for Video-and-Language Research". In: *International Conference on Computer Vision (ICCV)*. Oct. 2019. Oral.

[C39] Yao Fu, Hao Zhou, Jiaze Chen, **Lei Li**. "Rethinking Text Attribute Transfer: A Lexical Analysis". In: *the 12th International Conference on Natural Language Generation (INLG)*. Oct. 2019.

[C38] Yunfei Lu, Linyun Yu, Peng Cui, Chengxi Zang, Renzhe Xu, Yihao Liu, **Lei Li**, Wenwu Zhu. "Uncovering the Co-driven Mechanism of Social and Content Links in User Churn Phenomena". In: *the 25th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. New York, NY, USA: ACM, Aug. 2019.

[C37] Rongxiang Weng, Hao Zhou, Shujian Huang, **Lei Li**, Yifan Xia, Jiajun Chen. "Correct-and-Memorize: Learning to Translate from Interactive Revisions". In: *the 28th International Joint Conference on Artificial Intelligence (IJCAI)*. Aug. 2019. Oral.

[C36] Yu Bao, Hao Zhou, Shujian Huang, **Lei Li**, Lili Mou, Olga Vechtomova, Xinyu Dai, Jiajun Chen. "Generating Sentences from Disentangled Syntactic and Semantic Spaces". In: *the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*. July 2019.

[C35] Lin Qiu, Yunxuan Xiao, Yanru Qu, Hao Zhou, **Lei Li**, Weinan Zhang, Yong Yu. "Dynamically Fused Graph Network for Multi-hop Reasoning". In: *the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*. July 2019.

[C34] Huangzhao Zhang, Hao Zhou, Ning Miao, **Lei Li**. "Generating Fluent Adversarial Examples for Natural Languages". In: *the 57th Annual Meeting of the Association for Computational Linguistics (ACL) - short papers*. July 2019.

[C33] Hao Wu, Jiayuan Mao, Yufeng Zhang, Yuning Jiang, **Lei Li**, Weiwei Sun, Weiying Ma. "Unified Visual-Semantic Embeddings: Bridging Vision and Language with Structured Meaning Representations". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019. Oral.

[C32] Ning Miao, Hao Zhou, Lili Mou, Rui Yan, **Lei Li**. "CGMH: Constrained Sentence Generation by Metropolis-Hastings Sampling". In: *the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. Jan. 2019. Oral.

[C31] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Yitan Li, **Lei Li**. "BRITS: Bidirectional Recurrent Imputation for Time Series". In: *the 32nd Conference on Neural Information Processing Systems (NeurIPS)*. Dec. 2018.

[C30] Haoyue Shi, Hao Zhou, Jiaze Chen, **Lei Li**. "On Tree-Based Neural Sentence Modeling". In: *Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Oct. 2018.

[C29] **Lei Li**, Xiaojun Wan. "Overview of the NLPCC 2018 shared task: Single document summarization". In: *Proc. of NLPCC*. Oct. 2018.

[C28] Jiawei Wu, **Lei Li**, William Yang Wang. "Reinforced Co-Training". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. New Orleans, Louisiana: Association for Computational Linguistics, June 2018, pp. 1252–1262.

[C27] Lifeng Hua, Xiaojun Wan, **Lei Li**. "Overview of the NLPCC 2017 shared task: Single document summarization". In: *Proc. of NLPCC*. Oct. 2017.

[C26] Yusuf Erol, Yi Wu, **Lei Li**, Stuart Russell. "A Nearly-Black-Box Online Algorithm for Joint Parameter and State Estimation in Temporal Models". In: *the 31st AAAI Conference on Artificial Intelligence (AAAI)*. July 2017.

[C25] Yi Wu, **Lei Li**, Stuart Russell, Rastislav Bodik. "Swift: Compiled Inference for Probabilistic Programming Languages". In: *25th International Joint Conference on Artificial Intelligence (IJCAI)*. July 2016.

[C24] Zihang Dai, **Lei Li**, Wei Xu. "CFO: Conditional Focused Neural Question Answering with Large-scale Knowledge Bases". In: *the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*. July 2016.

[C23] Da-Cheng Juan, **Lei Li**, Huan-Kai Peng, Diana Marculescu, Christos Faloutsos. "Beyond Poisson: Modeling Inter-Arrival Times of Requests in a Datacenter". In: *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*. 2014.

[C22] **Lei Li**, Bharath Ramsundar, Stuart Russell. "Dynamic Scaled Sampling for Deterministic Constraints". In: *16th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2013.

[C21] Sharad Vikram, **Lei Li**, Stuart Russell. "Handwriting and Gestures in the Air, Recognizing on the Fly". In: *ACM Conference on Human Factors in Computing Systems (CHI) Extended Abstracts*. 2013.

[C20] Siyuan Liu, **Lei Li**, Ramayya Krishnan. "Hibernating Process: Modelling Mobile Calls at Multiple Scales". In: *IEEE International Conference on Data Mining (ICDM)*. 2013.

[C19] Yusuf Erol, **Lei Li**, Bharath Ramsundar, Stuart Russell. "The Extended Parameter Filter". In: *Proceedings of the 30th International Conference on Machine learning (ICML)*. 2013.

[C18] Bin Fu, Jialiu Lin, **Lei Li**, Christos Faloutsos, Jason Hong, Norman Sadeh. "Why People Hate Your App - Making Sense of User Feedback in a Mobile App Store". In: *the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. New York, NY, USA: ACM, 2013.

[C17] Mark Rogers, **Lei Li**, Stuart Russell. "Multilinear Dynamical Systems for Tensor Time Series". In: *the 27th Conference on Neural Information Processing Systems(NeurIPS)*. 2013.

[C16] Yasuko Matsubara, **Lei Li**, Evangelos E. Papalexakis, David Lo, Yasushi Sakurai, Christos Faloutsos. "F-Trail: Finding Patterns in Taxi Trajectories". In: *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*. 2013, pp. 86–98.

[C15] Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Hanghang Tong, Sugato Basu, Leman Akoglu, Danai Koutra, Christos Faloutsos, **Lei Li**. "RolX: Structural Role Extraction and Mining in Large Graphs". In: *Proceeding of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. New York, NY, USA: ACM, 2012.

[C14] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, **Lei Li**, Christos Faloutsos. "Rise and Fall Patterns of Information Diffusion: Model and Implications". In: *Proceeding of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. New York, NY, USA: ACM, 2012.

[C13] Keith Henderson, Brian Gallagher, **Lei Li**, Leman Akoglu, Tina Eliassi-Rad, Hanghang Tong, Christos Faloutsos. "It's Who You Know: Graph Mining Using Recursive Structural Features". In: *Proceeding of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. San Diego, California: ACM, 2011.

[C12] **Lei Li**, Chieh-Jan Mike Liang, Jie Liu, Suman Nath, Andreas Terzis, Christos Faloutsos. "ThermoCast: A Cyber-Physical Forecasting Model for Data Centers". In: *Proceeding of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. San Diego, California: ACM, 2011.

[C11] **Lei Li**, B. Aditya Prakash. "Time Series Clustering: Complex is Simpler!" In: *Proceedings of the 28th International Conference on Machine Learning (ICML)*. Bellevue, Washington, 2011.

[C10] Yasushi Sakurai, **Lei Li**, Yasuko Matsubara, Christos Faloutsos. "WindMine: Fast and Effective Mining of Web-click Sequences". In: *SIAM International Conference on Data Mining (SDM)*. 2011.

[C9] **Lei Li**, B. Aditya Prakash, Christos Faloutsos. "Parsimonious linear fingerprinting for time series". In: *the Proceedings of the Very Large Data Bases Endowment (VLDB)*. Vol. 3. VLDB Endowment, Sept. 2010, pp. 385–396.

[C8] Keith Henderson, Tina Eliassi-Rad, Christos Faloutsos, Leman Akoglu, **Lei Li**, Koji Maruhashi, B. Aditya Prakash, Hanghang Tong. "Metric forensics: a multi-level approach for mining volatile graphs". In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. KDD '10. Washington, DC, USA: ACM, 2010, pp. 163–172.

[C7] **Lei Li**, James McCann, Nancy Pollard, Christos Faloutsos. "BoLeRO: a principled technique for including bone length constraints in motion capture occlusion filling". In: *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*. SCA '10. Madrid, Spain: Eurographics Association, 2010, pp. 179–188.

[C6] **Lei Li**, James McCann, Nancy Pollard, Christos Faloutsos. "DynaMMo: Mining and Summarization of Coevolving Sequences with Missing Values". In: *Proceeding of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. Paris, France: ACM, 2009.

[C5] Fan Guo, **Lei Li**, Christos Faloutsos, Eric P. Xing. "C-DEM: a Multi-modal Query System for Drosophila Embryo Databases". In: *the Proceedings of the Very Large Data Bases Endowment (VLDB)*. Vol. 1. VLDB Endowment, Aug. 2008, pp. 1508–1511.

[C4] **Lei Li**, Wenjie Fu, Fan Guo, Todd C. Mowry, Christos Faloutsos. "Cut-and-Stitch: efficient parallel learning of linear dynamical systems on smps". In: *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*. Las Vegas, Nevada, USA: ACM, 2008, pp. 471–479.

[C3] Yasushi Sakurai, Rosalynn Chong, **Lei Li**, Christos Faloutsos. "Efficient Distribution Mining and Classification". In: *SIAM International Conference on Data Mining (SDM)*. 2008, pp. 632–643.

[C2] **Lei Li**, James McCann, Christos Faloutsos, Nancy Pollard. "Laziness is a virtue: Motion stitching using effort minimization". In: *The 29th Annual Conference of the European Association for Computer Graphics (EG), Short Paper Proceedings*. 2008.

[C1] **Lei Li**, Qiaoling Liu, Yunfeng Tao, Lei Zhang, Jian Zhou, Yong Yu. "Providing an Uncertainty Reasoning Service for Semantic Web Application". In: *Asia-Pacific Web Conference*. 2006, pp. 628–639.

Refereed Workshop Papers

[W15] Xi Xu, Siqi Ouyang, Brian Yan, Patrick Fernandes, William Chen, **Lei Li**, Graham Neubig, Shinji Watanabe. "CMU's IWSLT 2024 Simultaneous Speech Translation System". In: *Proceedings of the 21st International Conference on Spoken Language Translation (IWSLT 2024)*. Aug. 2024, pp. 154–159.

[W14] Brian Yan, Patrick Fernandes, Jinchuan Tian, Siqi Ouyang, William Chen, Karen Livescu, **Lei Li**, Graham Neubig, Shinji Watanabe. "CMU's IWSLT 2024 Offline Speech Translation System: A Cascaded Approach For Long-Form Robustness". In: *Proceedings of the 21st International Conference on Spoken Language Translation (IWSLT 2024)*. Association for Computational Linguistics, Aug. 2024, pp. 164–169.

[W13] Tianrui Gu, Kaie Chen, Siqi Ouyang, **Lei Li**. "PlayGround Low Resource Machine Translation System for the 2023 AmericasNLP Shared Task". In: *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*. Ed. by Manuel Mager, Abteen Ebrahimi, Arturo Oncevay, Enora Rice, Shruti Rijhwani, Alexis Palmer, and Katharina Kann. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 173–176.

[W12] Siqi Ouyang, Rong Ye, **Lei Li**. "On the Impact of Noises in Crowd-Sourced Data for Speech Translation". In: *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*. Association for Computational Linguistics, May 2022, pp. 92–97.

[W11] Lihua Qian, Yi Zhou, Zaixiang Zheng, Yaoming Zhu, Zehui Lin, Jiangtao Feng, Shanbo Cheng, **Lei Li**, Mingxuan Wang, Hao Zhou. "The Volctrans GLAT System: Non-autoregressive Translation Meets WMT21". In: *Sixth Conference on Machine Translation (WMT21)*. Nov. 2021.

[W10] Chengqi Zhao, Zhicheng Liu, Jian Tong, Tao Wang, Mingxuan Wang, Rong Ye, Qianqian Dong, Jun Cao, **Lei Li**. "The Volctrans Neural Speech Translation System for IWSLT 2021". In: *The International Conference on Spoken Language Translation (IWSLT)*. Aug. 2021.

[W9] Liwei Wu, Xiao Pan, Zehui Lin, Yaoming Zhu, Mingxuan Wang, **Lei Li**. "The Volctrans Machine Translation System for WMT20". In: *Proceedings of the Fifth Conference on Machine Translation (Volume 2: Shared Task Papers)*. Nov. 2020.

[W8] Runxin Xu, Zhuo Zhi, Jun Cao, Mingxuan Wang, **Lei Li**. "Volctrans Parallel Corpus Filtering System for WMT 2020". In: *Proceedings of the Fifth Conference on Machine Translation (Volume 2: Shared Task Papers)*. Nov. 2020.

[W7]   Gen Li, Shikun Xu, Xiang Liu, **Lei Li**, Changhu Wang. "Jersey Number Recognition with Semi-Supervised Spatial Transformer Network". In: *IEEE Conference on Computer Vision and Pattern Recognition workshops, Computer Vision in Sports*. June 2018, pp. 1864 –1871.

[W6]   Yi Wu, **Lei Li**, Stuart Russell. "BFiT: From Possible-World Semantics to Random-Evaluation Semantics in Open Universe". In: *Neural Information Processing Systems, Probabilistic Programming workshop*. 2014.

[W5]   Simon Shaolei Du, Yilin Liu, Boyi Chen, **Lei Li**. "Maxios: Large Scale Nonnegative Matrix Factorization for Collaborative Filtering". In: *Neural Information Processing Systems, workshop on Distributed Machine Learning and Matrix Computations*. 2014.

[W4]   Siyuan Liu, **Lei Li**, Christos Faloutsos, Lionel Ni. "Mobile Phone Graph Evolution: Findings, Model and Interpretation". In: *IEEE International Conference on Data Mining, workshop on Data Mining Technologies for Computational Collective Intelligence*. 2011.

[W3]   **Lei Li**. "Fast Algorithms for Time Series Mining". In: *26th IEEE International Conference on Data Engineering, PHD Workshop*. 2010, pp. 341–344.

[W2]   Fan Guo, **Lei Li**, Christos Faloutsos. "Tailoring click models to user goals". In: *Proceedings of the 2009 workshop on Web Search Click Data*. WSCD '09. Barcelona, Spain: ACM, 2009, pp. 88–92.

[W1]   Wanhong Xu, Xi Zhou, **Lei Li**. "Inferring privacy information via social relations". In: *IEEE 24th International Conference on Data Engineering workshops*. 2008, pp. 525–530.

Refereed System Demonstration Papers

[D5]   Chinmay Dandekar, Wenda Xu, Xi Xu, Siqi Ouyang, **Lei Li**. "Translation Canvas: An Explainable Interface to Pinpoint and Analyze Translation Systems". In: *the Conference on Empirical Methods in Natural Language Processing (EMNLP): System Demonstrations*. Nov. 2024.

[D4]   Chengqi Zhao, Mingxuan Wang, Qianqian Dong, Rong Ye, **Lei Li**. "NeurST: Neural Speech Translation Toolkit". In: *the 59th Annual Meeting of the Association for Computational Linguistics (ACL): System Demonstrations*. Aug. 2021.

[D3]   Dongyu Ru, Zhenghui Wang, Lin Qiu, Hao Zhou, **Lei Li**, Weinan Zhang, Yong Yu. "QuAChIE: Question Answering based Chinese Information Extraction System". In: *the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) - System Demonstrations*. July 2020.

[D2]   Runxin Xu, Jun Cao, Mingxuan Wang, Jiaze Chen, Hao Zhou, Ying Zeng, Yuping Wang, Li Chen, Xiang Yin, Xijin Zhang, Songcheng Jiang, Yuxuan Wang, **Lei Li**. "Xiaomingbot: A Multilingual Robot News Reporter". In: *the 58th Annual Meeting of the Association for Computational Linguistics (ACL): System Demonstrations*. July 2020.

[D1]   Zhaoyue Sun, Jiaze Chen, Hao Zhou, Deyu Zhou, **Lei Li**, Mingmin Jiang. "GraspSnooker: Automatic Chinese Commentary Generation for Snooker Videos". In: *the 28th International Joint Conference on Artificial Intelligence (IJCAI)*. Aug. 2019, pp. 6569–6571. Demo.

Technical Reports

[T2]   **Lei Li**, Stuart Russell. *The BLOG Language Reference*. Tech. rep. UCB/EECS-2013-51. EECS Department, University of California, Berkeley, May 2013.

[T1]   **Lei Li**. "Fast algorithms for mining co-evolving time series". PhD thesis. Carnegie Mellon University, 2011.

Patents and Invention Disclosures

[P42]   Jingjing Xu, Chun Gan, Hao Zhou, **Lei Li**, Zaixiang Zheng. "Vocabulary generation for neural machine translation". Pat. 12112139B2. Oct. 2024.

[P41]   **Lei Li**, Jun Cao, Minguxan Wang, Zhou Qian. "Method for translating words in a picture, electronic device, and storage medium". U.S. pat. 11,954,455 B2. Apr. 2024.

[P40] Gen Li, Yi He, **Lei Li**, Yitan Li. "Media retrieval method and apparatus". Pat. 11874869B2. Jan. 2024.

[P39] Yuzhang Du, Peihao Zhu, Yiming Chen, Chongxing Zhou, Mingxuan Wang, **Lei Li**. "Video translation method and apparatus, storage medium, and electronic device". Pat. 11580314B2. Sept. 2023.

[P38] **Lei Li**, Mingxuan Wang, Hao Zhou, Zewei Sun. "Interactive machine translation method, electronic device, and computer-readable storage medium". Pat. 11704504B2. July 2023.

[P37] Yuzhang Du, Peihao Zhu, Chongxing Zhou, Yiming Chen, Mingxuan Wang, **Lei Li**. "Document translation method and apparatus, storage medium, and electronic device". Pat. 11580314B2. Feb. 2023.

[P36] Yi He, **Lei Li**, Cheng Yang, Gen Li, Yitan Li. "Method and device for comparing media features". Pat. 11593582B2. Feb. 2023.

[P35] Mingxuan Wang, Qianqian Dong, **Lei Li**. "Speech translation method electronic device and computer-readable storage medium using SEQ2SEQ for determining alternative translated speech segments". Pat. 11,586,831 B2. Feb. 2023.

[P34] Yi He, **Lei Li**, Cheng Yang, Gen Li, Yitan Li. "Video Feature Extraction Method and Device". Pat. 11,455,802 B2. Sept. 2022.

[P33] Cheng Yang, Yi He, **Lei Li**. "Method and device for processing feature point of image". Pat. 11403835B2. Aug. 2022.

[P32] Jun Cao, **Lei Li**, Mingxuan Wang, Peihao Zhu. "Method for acquiring a parallel corpus, electronic device, and storage medium". Pat. 11379664B2. July 2022.

[P31] Yi He, **Lei Li**, Cheng Yang, Gen Li, Yitan Li. "Method and device for determining duplicate video". Pat. 11,265,598 B2. Mar. 2022.

[P30] Gen Li, **Lei Li**, Yi He. "Audio Retrieval and Identification Method and Device". Pat. 11,182,426 B2. Nov. 2021.

[P29] Yi He, **Lei Li**, Cheng Yang, Gen Li, Yitan Li. "Video Feature Extraction Method and Device". Pat. 11,055,536 B2. July 2021.

[P28] Gen Li, Shikun Xu, Yandong Zhu, **Lei Li**, Changhu Wang. "Target Object Image Detection Method and Device". Pat. 201811010095.X. May 2021.

[P27] Yangyu Chen, Yi He, **Lei Li**. "Method and Device for determining geometric transformation relation for images". Pat. 10,984,542 B2. Apr. 2021.

[P26] Yi He, Gen Li, **Lei Li**. "Image processing method and device". Pat. 201910498629.6. Apr. 2021.

[P25] Xunpeng Huang, Zhengyang Liu, **Lei Li**. "Target Object Classification method and device". Pat. 202010057296.6. Apr. 2021.

[P24] **Lei Li**, Jiaze Chen, Jiamin Chen, Weiying Ma, Lifeng Hua. "Method and Device for generating information". Pat. 201811455645.9. Apr. 2021.

[P23] Gen Li, **Lei Li**, Yi He. "Audio Fingerprint Extraction Method and Device". Pat. 10,950,255 B2. Mar. 2021.

[P22] Heng She, Yang Wang, Yinuo Guo, Huiru Zhang, Yitan Li, **Lei Li**, Hang Li. "Method and Device for Push-Notifying Information". Pat. 201811562666.0. Mar. 2021.

[P21] Jiaze Chen, **Lei Li**, Ying Zeng, Weiying Ma. "Method and Device for generation product description information". Pat. 201811457980.2. Jan. 2021.

[P20] Jiangdong Deng, **Lei Li**, Weiying Ma. "Method and device for stock selection". Pat. 201810910344.4. Jan. 2021.

[P19] Yi He, Cheng Yang, Gen Li, Yitan Li, **Lei Li**. "Duplicate video detection method and device". Pat. 201810273706.3. Jan. 2021.

[P18] Jiangdong Deng, **Lei Li**, Weiying Ma. "Sentiment Prediction Method and Device". Pat. 201810909879.X. Sept. 2020.

[P17] Yi He, **Lei Li**, Cheng Yang, Gen Li, Yitan Li. "A Method and Device for Video Feature Extraction". Pat. 201810271774.6. July 2020.

[P16] Hao Zhou, **Lei Li**. "Machine Translation Method and Device". Pat. 201910105606.4. July 2020.

[P15] Yi He, **Lei Li**, Cheng Yang, Gen Li, Yitan Li. "Video Feature Extraction Method and Device". Pat. 201810271773.1. May 2020.

[P14] Hao Zhou, **Lei Li**, Ning Miao. "Method and Device for Generating Text". Pat. 201910105002.X. May 2020.

[P13] Gen Li, Shikun Xu, Yandong Zhu, **Lei Li**, Changhu Wang. "Target Object Image Detection Method and Device". Pat. 201811010092.6. Apr. 2020.

[P12] Linyun Yu, **Lei Li**, Haibin Yin, Wenjia Zhu, Dong Jiang. "Method and Apparatus for generating image". Pat. 201810668219.7. Apr. 2020.

[P11] Jiangdong Deng, Qu Peng, **Lei Li**, Weiying Ma. "A method for outputing information". Pat. 201811074033.5. Mar. 2020.

[P10] **Lei Li**, Zihang Dai, Wei Xu. "Systems and methods for human inspired simple question answering (HISQA)". Pat. 10606846B2. Mar. 2020.

[P9] Hao Zhou, **Lei Li**, Jiaze Chen, Haoyue Shi. "Method and Device for generating information". Pat. 201910105241.5. Mar. 2020.

[P8] Hao Zhou, **Lei Li**. "Method and Device for generating information". Pat. 201910105235.X. Mar. 2020.

[P7] Yi He, **Lei Li**, Xianzi Zong, Hao Tang, Guangguo Zheng. "Method and Device for searching information". Pat. 201811060981.3. Jan. 2020.

[P6] Linyun Yu, **Lei Li**, Haibin Yin, Dong Jiang. "Method and apparatus for generating image". Pat. 201810669838.8. Jan. 2020.

[P5] Yangyu Chen, Yi He, **Lei Li**. "A method and apparatus for determining a geometric transformation relationship between images". Pat. 201811060837.X. Nov. 2019.

[P4] Dong Jiang, Yanbin Zhao, Shuang Hou, Xuhong Xia, **Lei Li**, Dingkun Hong. "Construction method and device of voice classification model." Pat. 201710388497.2. Mar. 2019.

[P3] Zheng Chen, **Lei Li**, Chenxi Lin, Qiaoling Liu, Jian Wang, Benyu Zhang. "Adaptive grouping in a file network". Pat. 7,634,471. Dec. 2009.

[P2] Zheng Chen, **Lei Li**, Chenxi Lin, Qiaoling Liu, Jian Wang, Benyu Zhang. "System and method for exploring a semantic file network". Pat. 7,624,130. Nov. 2009.

[P1] Zheng Chen, **Lei Li**, Chenxi Lin, Qiaoling Liu, Jian Wang, Benyu Zhang. "Extracting Semantic Attributes". Pat. 7,502,785. Mar. 2009.

Software Artifacts

[S5] *BLOG: A Probabilistic Modeling System for Bayesian Logic Probabilistic Programming Language*. Java. Latest version (v0.10). Brian Milch was the author for version 0.3 or earlier. URL: `http://bayesianlogic.github.io/`.

[S4] *DynaMMo: a machine learning toolbox for multi-dimensional co-evolving time series*. Matlab. URL: `https://github.com/lileicc/dynammo`.

[S3] *LightSeq: A High Performance GPU Training and Inference Library for Transformer Models*. URL: `https://github.com/bytedance/lightseq`.

[S2] *Swift: A compiler for BLOG probabilistic programming language*. URL: `https://github.com/lileicc/swift`.

[S1] *VolcTrans: a machine translation platform for 100+ languages*. URL: `https://translate.volcengine.com`.

Talks

1. Invited Talk at Bosch US Lab: Scalable Post-Training Optimization for Large Language Models. May 23, 2025.
2. Invited Talk at NEC Lab America: Scalable Post-Training Optimization for Large Language Models. March 6, 2025. Princeton, NJ.
3. Princeton University AI2 Distinguished Lecture: Generative AI for Functional Protein Design. March 5, 2025. Princeton, NJ.
4. CMU-Pitt CPCB Seminar: Generative AI for Functional Protein Design. Jan 24, 2025.
5. NeurIPS 2024 Tutorial: Watermarking for Large Language Models. December 10, 2024 in Vancouver, Canada. (with Xuandong Zhao and Yu-Xiang Wang)
6. CMU FLAME Talk: The Science of Evaluation and Alignment for Large Language Models. December 6, 2024.
7. Invited Talk at University of Massachusetts Boston: Towards Scaling Large Language Models to 1000 Languages - Challenges and Advances. Nov 4, 2024. Boston, MA.
8. Invited Talk at Northeastern University: The Science of Evaluation and Alignment for Large Language Models. November 4, 2024. Boston, MA.
9. Invited Talk at MIT (FutureTech workshop on Role of AI in Science): Watermarking and Detecting AI Generation. November 3, 2024. Boston, MA.
10. University of Pittsburgh Intelligent Systems Program Forum talk: The Science of Evaluation and Alignment for Large Language Models. October 25, 2024. Pittsburgh, PA.
11. CMU LTI Colloquium: From Words to Molecules: Harnessing Generative AI for Breakthroughs in Language and Molecular Design. September 13, 2024. Pittsburgh, PA.
12. Invited Talk at Genentech: Generative AI for Functional Protein Design. September 10, 2024, in South San Francisco, California.
13. Keynote at the Language+Molecule workshop at ACL 2024: A Tale of Two Realms: Commonalities and Distinctions of Generating Language and Molecules. August 15, 2024 in Bangkok, Thailand.
14. ACL 2024 Tutorial: Watermarking for Large Language Models. August 11, 2024 in Bangkok, Thailand. (with Xuandong Zhao and Yu-Xiang Wang)
15. NAACL 2024 Tutorial: Combating Security and Privacy Issues in the Era of Large Language Models. June 16, 2024 in Mexico City, Mexico. (with Muhao Chen, Chaowei Xiao, Huan Sun, and Leon Derczynski).
16. CMU 15641 Machine Learning System Guest Lecture: Assessing and Improving Large Language Models. April 10, 2024.
17. Michigan State University CSE 891 LLM Guest Lecture: Assessing and Improving Large Language Models. April 1, 2024.
18. CMU CyLab Seminar: Is it generated by AI? Attacks and Robust Watermarking for Generative AI. January 22, 2024.
19. Ohio State University TDAI's Foundations of Data Science and AI Speaker: Self-assisting and Cooperative Large Language Models. January 12, 2024.
20. Nanyang Technological University Invited Talk: Empowering Responsible Use of Large Language Models. Singapore. December 11, 2023.
21. National University of Singapore: Self-assisting and Cooperative Large Language Models. December 8, 2023.
22. CMU LTI Colloquium: Empowering Responsible Use of Large Language Models. September 8, 2023.
23. Johns Hopkins University CLSP Seminar: Empowering Responsible Use of Large Language Models. Baltimore, MD. September 1, 2023.
24. Amazon AWS AI Invited Talk: Learning When to Translate for Streaming Speech. Lego-MT: Learning Detachable Models for Massive Multilingual Machine Translation. Santa Clara, CA. August 4, 2023.
25. CASIS Invited Talk: Trustworthy Large Language Models. 27th Annual Workshop of LLNL Center for Advanced Signal and Imaging Sciences. August 3, 2023.

26. Guest Lecture: Breaking Language Barriers with Massive Multilingual Machine Translation. IIIS at Tsinghua University. May 23, 2023.

27. UCSD CS Invited Talk: Breaking Language Barriers with Massive Multilingual Machine Translation. May 12, 2023.

28. UCLA CS Invited Talk: Breaking Language Barriers with Massive Multilingual Machine Translation. May 11, 2023.

29. UCSB PSTAT Seminar: Breaking Language Barriers with Massive Multilingual Machine Translation. April 12, 2023.

30. Invited Talk at Pepperdine University: Breaking Language Barriers with Massive Multilingual Machine Translation. March 21, 2023.

31. Invited Talk: Right for the Right Reasons: Explainable Neural Reasoning, Shanghai Jiao Tong University, Nov. 8, 2022.

32. Talk: Deep Generative Models for Drug Discovery, Responsible Machine Learning Summit, UCSB, Oct 7, 2022.

33. CCF-ADL Tutorial: Pre-training for Neural Machine Translation. August 18, 2022 (online).

34. Invited Talk: Breaking the Language Barriers with Neural Machine Translation. IEEE Central Coast. Santa Barbara, August 17, 2022.

35. Invited Talk: Multi-objective Drug Molecule Generation, the AI for Health Forum of BAAI Conference. June 2, 2022.

36. Keynote: Challenges and Progress of Massive Multilingual Machine Translation. Pre-ACL 2022 Online Meetup. May 20, 2022.

37. Invited Talk: Challenges and Progress of Multilingual Machine Translation. Meta (Facebook) AI. April 15, 2022.

38. Invited Talk: Training Privacy-preserving Language Models. at International Conference on the next-generation AI risk governance and policies. February 18, 2022.

39. Invited Talk: Efficient Machine Translation. at Apple Inc. December 15, 2021.

40. Keynote: Efficient Machine Translation. at the 17th China Conference on Machine Translation. Xining, Qinghai, China. October 10, 2021.

41. ACL 2021 Tutorial: Pre-training for Neural Machine Translation (online), August 1, 2021.

42. Invited talk: Speech Translation – Streamlining Research and Product Innovation. the NLP forum of 2021 Global AI Technology Conference, Zhejiang Hangzhou, June 6, 2021.

43. Guest Lecture: Data mining for Personalized Recommendation Systems, undergraduate Data Mining class, Tsinghua University. June 4, 2021.

44. Talk: Interpretable, Controllable, and Scalable Natural Language Generation, the Natural Language Generation workshop (online), May 28, 2021.

45. Guest Lecture: Deep Generative Models for Text and Beyond. Deep Learning course. Department of Computer Science, Tsinghua University. April 28, 2021.

46. Guest Lecture: Deep Generative Models for Text and Beyond. Advanced Machine Learning course. Department of Computer Science, Tsinghua University. April 14, 2021.

47. Talk: Recent Advances in Machine Writing and Translation – Algorithms and Challenges, GTC - China, December 19, 2020.

48. Invited Talk: Controllable and Interpretable Machine Learning for Natural Language Generation. Department of Computer Science. Nanjing University. December 11, 2020.

49. Guest Lecture: Deep Latent Models for Natural Language Generation. Graduate course in Machine Learning, Department of Computer Science, Tsinghua University. December 8, 2020.

50. Invited Talk: Is it a fact? – Logic Enhanced Neural Reasoning. Cognitive Intelligence workshop, Shanghai, December 8, 2020.

51. Keynote Talk: Scalable, Controllable, and Interpretable Machine Learning for Natural Language Generation. NeurIPS 2020 Meetups in Beijing. December 6, 2020.

52. Guest Lecture: Scalable, Controllable and Interpretable Machine Learning for Natural Language Generation. undergraduate class in CS, IIIS, Tsinghua University. October 8, 2020.

53. Invited Talk: Constrained Text Generation – Monte Carlo Meets Neural Nets. Tsinghua University, Institute of Inter-disciplinary Information Sciences. October 8, 2020.

54. Guest Lecture: Learning Deep Latent Models for Text Sequences, graduate class in Machine Learning, Tsinghua University, April 29, 2020.

55. Talk at ICLR 2020: Learning Deep Latent Models for Text Sequences. April 29, 2020.

56. Guest Lecture, Deep Generative Models for Text Generation, graduate class in Machine Learning, Tsinghua University, November 26, 2019.

57. Invited Talk: Deep Text Generation: Models and Applications, School of Data Science and Engineering, East China Normal University, Shanghai, China, November 8, 2019.

58. Industrial Invited Talk at CIKM 2019: Multi-lingual cross-modal machine translation at global scale, Beijing, China, November 7, 2019.

59. EMNLP 2019 Tutorial: Discreteness in Natural Language Processing, Hong Kong, November 4, 2019. (3 hr, with Lili Mou and Hao Zhou)

60. CCF-ADL107/NLPCC 2019 Tutorial: Deep Generative Models for Text Generation, Dunhuang, China, Oct 9, 2019. (3 hr, with Hao Zhou).

61. Invited Talk at IJCAI 2019 Industrial Days: Deep Text Generation: Models and Applications, Macau, China, August 2019.

62. Invited Talk: Machine Writing and Constrained Text Generation, CCF-AI workshop at Beijing Jiao Tong University, June 2019.

63. Guest Lecture, Deep Natural Language Generation, graduate class in Machine Learning, Tsinghua University, May 15, 2019.

64. Guest Lecture, Deep Natural Language Generation, graduate class in NLP, Peking University, May 8, 2019.

65. Guest Lecture, Personalized Recommendation Systems in Practice, undergrad Data Mining class, Tsinghua University. April 26, 2019.

66. Invited Talk: Machine Writing and Constrained Sequence Generation, Arizona State University, Tempe, AZ, April 2019.

67. Invited Talk, Constrained Sentence Generation, School of Data Science, Fudan University, Shanghai, March 2019.

68. Invited Talk: Recent Progress in Natural Language Representation Learning and Generation, InfoQ AICon, Beijing, December 21, 2018.

69. Invited Talk: Learning to generate better news, translation, and advertisement. CSDN AI NEXTCon, Beijing, November 13, 2018.

70. Invited Talk: AI for Content Creation and Interaction, Department of Computer Science, University of Illinois Urbana Champaign, November 2018.

71. Guest Lecture, Deep Learning for Language Understanding and Generation, graduate Machine Learning class, Tsinghua University. November 13, 2018.

72. Invited Talk: AI for Content Creation and Interaction, Viterbi School of Engineering, University of Southern California, October 2018.

73. Invited Talk: Incubating Innovative IT Talents - an academic-industry fused perspective. Computer Education Conference, CCF, Hefei, China, July 2018.

74. Guest Lecture, Deep Learning for Language Understanding and Generation, graduate Machine Learning class, Tsinghua University. April 25, 2018.

75. Invited Talk: AI for Content Creation and Discovery, Computer Science Department, University of California at Davis, California, April 2018.

76. Invited Talk: Machine Writing and Language Generation, Computer Science Department, Fudan University, March 2018.

77. Invited Talk: AI for Content Creation and Interaction, ShanghaiTech University, March 2018

78. Invited Talk: AI for Content Creation and Interaction, Southeast University, March 2018

79. Invited Talk: AI for Content Creation and Interaction, Nanjing University, March 2018

80. Invited Talk: AI for Content Creation and Interaction, Nanjing Normal University, March 2018

81. Invited Talk: AI-assisted Content Creation, CAAI MIFS (Machine Intelligence Frontier Symposium), Beijing, October 2017.

82. Invited Talk: AI-Assisted Content Creation and Commenting, Institute of Computing at China Academy of Science, Beijing, September 2017.

83. Invited Talk: AI-powered Information Creation, Distribution, and Interaction, AI Computing Conference, Beijing, September 2017.

84. Invited Talk: AI Research at Toutiao, Data Science in China, KDD, Halifax, August 2017.

85. Guest Lecture, Deep Learning for Question Answering, graduate Machine Learning class, Tsinghua University. March 13, 2017.

86. CCF-ADL/NLPCC 2016 Tutorial: Deep Learning for Answering Questions, Kunming, China, December 2016.

87. Invited Talk: Learning to Understand and Generate Language, AIWorld Conference, Beijing, October 2016.

88. Invited Talk: Status and Challenges towards general AI. China Conference on Artificial Intelligence. Beijing, August 26-27, 2016.

89. Invited Talk: AI for Content Creation and Sharing, CCF-GAIR (Global Conference on AI and Robotics), Shenzhen, August 2016.

90. Invited Talk: Learning for Dialog and Question Answering, Shanghai Jiao Tong University, Shanghai, June 2016.

91. Invited Talk: Large-scale Learning for Language Understanding, University of Science and Technology in China, Hefei, May 2016.

92. Invited Talk: Learning for Language and Visual Understanding, CAAI-GAITC (Global AI Technology Conference), Beijing, April 2016.

93. Invited Keynote: Deep Learning – Towards More Intelligent Machines, China Computer Federation Young Computer Scientists and Engineers Forum, April 2015.

94. BLOG language and compiled inference. Computer Science department, Stanford University, and separately at Tsinghua University 2015.

95. BLOG language and efficient inference. Berkeley Programming Language Retreat, Santa Cruz, 2014

96. Scalable Probabilistic Inference for Complex Dynamical Models. Invited talk at Ohio State University, New York University, Emory University, Google (Mountain View), 2014.

97. Invited Fish Bowl Seminar: Fast Algorithms for Mining Co-evolving Time Series. , Texas A&M University, 2013.

98. DBLOG: an Open Universe Probabilistic Programming Language for Relational Modeling, The Second EITA Young Investigator Conference, Palo Alto, CA, July 2012.

99. Forecasting with Cyber-physical Interactions in Data Centers. CMU Parallel Data Lab. September 2011.

100. Proactive Detection of Insider Threats with Graph Analysis and Learning (CMU representative), DARPA ADMAS program kickoff conference, Arlington, VA. June 2011.

101. Fast Algorithms for Mining Co-evolving Time Series. Case Western Reserve University, PARC, Yahoo!, SIAT 2011.

102. KDD 2010 Tutorial: Indexing and Mining Time Sequences. Washington DC, KDD 2010. (with Christos Faloutsos)

103. Parsimonious Linear Fingerprinting for Time Series. CMU machine learning lunch, 2010.

104. PLiF: Parsimonious Linear Fingerprinting for Time Series. Zhejiang University, China. September 2010.

105. Fast Algorithms for Mining Co-evolving Time Series. SMU, NUS, HKUST, SJTU, December 2009.

106. Mining with Missing Values in Coevolving Time Series. MSR Redmond, May 2009.

107. Efficient Parallel Learning of Linear Dynamical Systems on SMPs. CMU machine learning lunch, 2008.

Teaching Experience

1. CMU 11781/02741: Generative AI for Biomedicine, Fall 2025.
2. CMU 11968: Large Language Model Systems (online), Fall 2025.
3. ACDL 2025: 5 lectures on Advanced Courses on Data Science and Machine Learning, June 9-13, 2025. Tuscany, Italy.
4. CMU 11868: Large Language Model Systems, Spring 2025.
5. CMU 11868: Large Language Model Systems, Spring 2024.
6. CMU 11737: Multilingual Natural Language Processing, Fall 2023.
7. UCSB CS190I: Deep Learning, undergraduate, Winter 2023.
8. UCSB CS291K: Machine Learning, graduate, Fall 2022.
9. UCSB CS165B: Machine Learning (focus on Deep Learning), undergraduate, Winter 2022.
10. UCSB CS291K: Deep Learning for Machine Translation, graduate, Fall 2021.
11. Lecturer in Probabilistic Programming for Advanced Machine Learning Summer School, Portland, OR, USA, 2014. (total 12hrs lecture for two weeks)
12. Teaching Assistant for
    10701 Machine Learning, instructor Ziv Bar-Joseph, CMU, Spring 2009.
    15415 Database Applications, instructor Christos Faloutsos, CMU, Fall 2008.
    Modern DBMS Implementation, SJTU, Fall 2005.
    Operating System Project(nachos), SJTU, Fall 2005.
    Modern Compiler Implementation, SJTU, Spring 2005.

Students Advising

Current Graduate Students:
1. Ramith Hettiarachchi (PhD student in CMU CPCB, 2024 -)
2. Andre Vicente Duarte (PhD student in CMU-Portugal Program, 2024 -)
3. Siqi Ouyang (PhD student at CMU, 2021-2023. 2024 -)
4. Zhenqiao Song (PhD student at CMU, 2022 - 2023. 2024 -)
5. Danqing Wang (PhD student at CMU, 2022 - 2023. 2024 -)
6. Kexun Zhang (PhD student at CMU, co-advised w/ William Wang, 2022 - 2023. 2024-)
7. Kaavya Subramanian (CMU LTI master student, capstone project, 2025.2 -)
8. Emily Shen (CMU LTI master student, capstone project, 2025.2 -)
9. Dhruv Garg (CMU LTI master student, capstone project, 2025.2 -)
10. Tim Han (CMU LTI master student, capstone project, 2025.2 -)
11. Venu Arvind Arangarajan (CMU LTI master student, capstone project, 2025.2 -)
12. Ziming Wang (CMU LTI master student, capstone project, 2025.2 -)
13. Arihant Sheth (CMU LTI master student, capstone project, 2025.2 -)
14. Yanlin Fei (CMU LTI master student, capstone project, 2025.1 -)
15. Sudeshna Merugu (Master, CMU LTI, 2024.12 -)
16. Yixiao (Jack) Zeng (Master, CMU LTI, 2024.9 -)
17. Xi Xu (Master, CMU LTI, 2023 -)
18. Brian Tufts (Master, CMU LTI , 2023 -)
19. Yee Man (Kath) Choi (Master, CMU LTI, 2023 -)

Past PhD Students:
1. Wenda Xu (UCSB PhD, co-advised w/ William Wang, 2021 - 2025.3, currently at Google Translation Research.)
2. Jielin Qiu (CMU PhD, graduated in 2024, currently at Boson.AI)
3. Xuandong Zhao (UCSB PhD student, co-advised w/ Yu-Xiang Wang, graduated in 2024, currently Postdoc researcher at UC Berkeley)

Past Master Students:
1. Zihan Ma (Master student at UCSB, 2022 - 2024)
2. Vihaan Rajendiran (Master student at UCSB, 2022 - 2024)

Thesis Committee
1. Bowen Tan (CMU PhD), proposal 2025.5.
2. Zhiqing Sun (CMU PhD), defense 2025.2.
3. Ta-Chung Chi (CMU PhD), defense 2024.
4. Weixi Feng (UCSB PhD), MAE 2023, Proposal 2024.
5. Kha-Dinh Luong (UCSB PhD), MAE 2023, Proposal 2024.
6. Qi Zeng (UIUC PhD), Defense 2023.
7. Yingrui Yang (UCSB PhD), Defense 2024.
8. Tsu-Jui Fu (UCSB PhD), MAE 2022, Proposal 2023, Defense 2024.
9. Xinlu Zhang (UCSB PhD), MAE 2022, Proposal 2023, Defense 2024.
10. Shiyang Li (UCSB PhD), MAE, Proposal 2022, Defense 2023.
11. Yuqing Zhu (UCSB PhD), MAE, 2022, Defense 2023.
12. Apoorva Jakalannanavar (UCSB, master thesis), 2022.

Other Graduate Mentees:
1. Songwen Zhao (visiting master student from Columbia University, 2025.1 -)
2. Xiwei Cheng (visiting master student from UC San Diego, June 2024 - Feb 2025)
3. Yuanjing Wei (Master, CMU LTI, Spring 2024)
4. Yujia Gao (Master, CMU LTI, Spring 2024)
5. Srinivas Gowriraj (Master, CMU LTI, Fall 2023)
6. Guanglei Zhu (Master, CMU RI, 2023 - 2024)
7. Qingxiu Dong (Jan 2022 - June 2023)
8. Jiangjie Chen (intern, April 2020 - May 2023)
9. Hong Zhao (intern at Bytedance, joint w/ Yuwei Yang, May 2021 - July 2021)
10. Peiyuan Zhang (intern at Bytedance, joint w/ Xunpeng Huang, Oct 2020 - May 2021)
11. Zaixiang Zheng (intern at Bytedance, joint w/ Hao Zhou, Sep 2020 - June 2021, currently researcher at Bytedance, MLNLC group)
12. Yaoming Zhu (intern at Bytedance, joint w/ Mingxuan Wang, 2020 - Mar. 2021, currently ML engineer at Bytedance, MLNLC group)
13. Lihua Qian (intern at Bytedance, joint w/ Hao Zhou, 2020 - Mar. 2021, currently ML engineer at Bytedance, MLNLC group)
14. Zehui Lin (intern at Bytedance, joint w/ Mingxuan Wang, April 2020 - Jan 2021, now ML engineer at ByteDance, MLNLC group)
15. Ya Jing (intern at Bytedance, joint w/ Tao Kong, Jan. 2020 - Jun. 2021, currently researcher at Bytednace, MLNLC group)
16. Mingxuan Jing (intern at Bytedance, joint w/ Tao Kong, Oct 2019 - Jun. 2021, currently researcher at CAS Institute of Software)

17. Jingjing Gong (intern at ByteDance, joint w/ Hao Zhou, July 2019 - May 2020)
18. Xinlong Wang (intern at Bytedance, joint w/ Tao Kong, July 2019 - May 2021, now researcher at BAAI)
19. Zewei Sun (phd s joint w/ Mingxuan Wang, June 2019 - June 2020, now ML engineer at ByteDance MLNLC group)
20. Yuxuan Song (intern at Bytedance, joint w/ Hao Zhou, June 2019 - April 2020, was software engineer at ByteDance, MLNLC group, now PhD student at Tsinghua Univ.)
21. Rong Ye (intern at Bytedance, joint w/ Hao Zhou, April 2019 - June 2020, now software engineer at ByteDance, MLNLC group)
22. Chi Sun (intern at Bytedance, joint w/ Hao Zhou, April 2019 - Dec 2019)
23. Lin Qiu (intern at Bytedance, joint w/ Hao Zhou, Jan 2019 - Dec 2019, now at Amazon Shanghai)
24. Yu Bao (intern at Bytedance, joint w/ Hao Zhou, Jan 2019 - Nov 2019, now researcher at Bytedance)
25. Qianqian Dong (intern at ByteDance, joint w/ Mingxuan Wang, Nov. 2018 - Jun. 2021, now researcher at ByteDance MLNLC group)
26. Freda(Haoyue) Shi (intern at Bytedance, joint w/ Hao Zhou, PKU undergrad, now Ph.D. student at TTIC)
27. Yunfei Lu (intern, 2019, now Ph.D. student at Tsinghua Univ.)
28. Wenxian Shi (intern at Bytedance April 2018 - June 2019, joint w/ Hao Zhou, was ML engineer at Bytedance, MLNLC group, now PhD. student at MIT)
29. Ning Miao (intern at Bytedance March 2018 - June 2019, joint w/ Hao Zhou, was ML engineer at Bytedance, now Ph.D. student at Oxford Univ.)
30. Bohan Li (intern at Bytedance 2017, and Jan 2020 - May 2020, currently SDE at Bytedance.)
31. Ying Zeng (intern at Bytedance Oct 2017 - June 2018, currently SDE at Bytedance, MLNLC group)
32. Chengqi Zhao (intern at Bytedance June 2017 - September 2017, currently SDE at Bytedance, MLNLC group)
33. Gen Li (intern at Bytedance April 2016 - December 2016, currently SDE at Bytedance)
34. Yi Wu (intern at UC Berkeley, Aug 2013 -2014 (joint w/ Stuart Russell), summer intern 2016 at Bytedance, currently assistant professor at Tsinghua)
35. Fei Jiang (intern at Baidu, master from Tsinghua, April 2015 - Oct 2015, now at JP Morgan Chase)
36. Zihang Dai (Intern at Baidu, CMU, June 2015 - Aug 2015, now Research Scientist at x.AI)
37. Hieu Pham (intern at Baidu, Stanford master, June 2015 - Sep 2015, now Research Scientist at Google)
38. Di Zhou (intern at Baidu, from U. Texas, May 2015 - Aug 2015)
39. Huanhuan Wang (intern at CMU, joint w/ Siyuan Liu, sp 2012, currently at Google)
40. Shenghao Huang (intern at CMU, joint w/ Siyuan Liu, sp 2012, currently at LinkedIn)
41. Mark Rogers (intern at UC Berkeley, Jan 2012 - Dec 2013, joint w/ Stuart Russell)

Undergraduate Mentees
1. Daniel Ji (CMU, 2025.3 -)
2. Angela Chen (CMU, 2025.2 -)
3. Kadin Zhang (CMU, 2025.1 -)
4. Aidan Zhang (CMU, 2024.9 -)
5. Zhongmou He (visiting student from University of Michigan, undergrad, May 2024 -)
6. Daniel Kathein (CMU, May 2024 - Aug 2024)
7. Ally Du (CMU, May 2024 - Nov 2024)
8. Tianrui Gu (UCSB, 2023)
9. Kaie Chen (UCSB, 2023)
10. Jenny Wu (UCSB, 2023)
11. Chinmay Dandekar (UCSB, 2023 - 2024)
12. Yufei Song (UCSB, 2023)

13. Isaac Sloan (UC Leeds Scholar, June 2022 - Aug 2022)
14. Saisidharth Majeti (UCSB, 2022 - 2023)
15. Jamie Wang (UCSB, 2022)
16. Ahmed Elkordy (UCSB, 2022)
17. Yunfei Li (Jul. 2020 - Aug 2021, joint w/ Tao Kong, currently Ph.D. student at Tsinghua)
18. Meihua Dang (intern at Bytedance, Jan 2021 - Aug 2021, currently master student at UCLA CS)
19. Jiarui Lu (joint w/ Yuwei Yang, Dec. 2020 - July 2021, current at Mila)
20. Quanyu Long (joint w/ Mingxuan Wang, now Ph.D. student at NTU Singapore)
21. Chi Han (joint w/ Mingxuan Wang, July 2020 - July 2021, currently Ph.D. student at UIUC)
22. Yutong Xie (joint w/ Hao Zhou, Feb 2020 - Jan 2021, currently PhD student at UMich)
23. Qi Zeng (intern at Bytedance, summer 2019, joint w/ Hao Zhou, currently research scientist at Meta)
24. Danlu Chen (intern at Bytedance, joint w/ Hao Zhou, Oct 2019 - Jan 2020, now Ph.D. student at UCSD)
25. Wenxiao Wang (intern at Bytedance May - Nov 2018, currently PhD student at UMD)
26. Yao Fu (intern at Bytedance Dec. 2017 - Jun. 2018, now Ph.D. student at University of Edinburgh)
27. Shiyi Han (intern at Bytedance, now software engineer at Apple.)
28. Freda(Haoyue) Shi (intern at Bytedance, joint w/ Hao Zhou, PKU undergrad, now assistant professor at University of Waterloo)
29. Zefu Lu (Intern at Baidu, UIUC undergrad, May 2015 - Aug 2015)
30. Di Zhou (intern at Baidu, from U. Texas, May 2015 - Aug 2015)
31. Da Tang (intern at UC Berkeley, July 2014 - Nov 2014, currently Research Scientist at ByteDance.)
32. Lea Jia (intern at UC Berkeley, Feb 2014 - May 2014, Google)
33. Daniel Liu (intern at UC Berkeley, Feb 2014 - Oct 2014)
34. Tejas Srinivasan (intern at UC Berkeley, Oct 2013 - May 2014)
35. Boyi Chen (intern at UC Berkeley, Sep 2013 - Oct 2014, currently software engineer at LinkedIn)
36. Christopher Xie (intern at UC Berkeley, Aug 2013 - May 2014 currently at Meta Reality Lab)
37. Simon Du (intern at UC Berkeley, Jul 2013 - Jan 2015, currently assistant professor at U. Washington)
38. Zhiting Hu (intern at CMU, Jul 2013 - Aug 2013, currently assistant professor at UCSD)
39. Hao Fu (intern at CMU, Feb 2013 - May 2013)
40. Yiyao Jiang (intern at CMU, Feb 2013 - May 2013)
41. Bharath Ramsundar (intern at UC Berkeley, Jun 2012 - Dec 2013, founder and CEO of Deep Forest Sciences)
42. Gerald Fong (intern at UC Berkeley, Sep 2012 - May 2013, currently at Instagram)
43. Sharad Vikram (intern at UC Berkeley, Aug 2012 - May 2013, currently at Google)
44. Aaron Wong (intern at UC Berkeley, Jul 2012 - June 2013, current software engineer at Google)
45. Cheng Lai Victor Huang (intern at UC Berkeley, Sep 2012 - Jan 2013, Amazon)
46. Akihiro Matsukawa (intern at UC Berkeley, Oct 2011 - Jan 2012, joint w/ Stuart Russell, currently at D.E. Shaw)

K-12 Students
1. Harry Chen (Northwood High School, Irving, CA, 2023-2024. now at Caltech.)

Professional Service

Associate Editor/Action Editor:
1. Transactions on Machine Learning Research (TMLR), 2023 - .
2. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2021 - .

Conference/Workshop Organization

1. Organizer of ICML 2025 workshop on Generative AI for Biology.
2. NeurIPS 2024 Competition: Erasing the Invisible: A Stress-Test Challenge for Image Watermarks.
3. AAAI Doctoral Consortium Co-Chair 2022, 2023.
4. CCAI (China Conference on Artificial Intelligence): Co-Chair for Young Scientist Forum 2017, Chair for Forum on Innovative Applications of AI 2019,
5. CAAI-IEEE MIFS (Machine Intelligence Frontier Symposium) Co-Chair, 2016.
6. CAAI Byte Tech AI Symposium, PC Co-Chair, Beijing, January 2019.
7. KDD Cup Co-Chair 2017, Hands-on Tutorial Co-Chair 2018, 2023, Sponsorship Co-Chair 2019, 2020.
8. Organizer of Neurips 2023 workshop on Generative AI for Biology.
9. WSDM Cup Co-Chair 2022.
10. Organizer for 2018 Byte Cup International Machine Learning Competition. Task: News Title Generation. 1072 teams with 2175 participants from 35 countries. (Hosted by CAAI, ByteDance, & Tsinghua Knowledge Engineering Lab), link
11. Organizer for 2016 Byte Cup International Machine Learning Competition. Task: Expert Recommendation for Community Questions. 1000 teams with 1600 participants. (Hosted by CAAI, ByteDance, & Tsinghua Knowledge Engineering Lab), link

Conference Area Chair/Senior PC:

1. AAAI SPC 2019, 2020, 2023.
2. AACL Area Chair 2020. Senior Area Chair 2023.
3. ACL Senior Area Chair 2023.
4. ACL ARR Area Chair 2025.
5. CIKM Industrial Track SPC 2019.
6. COLM Area Chair 2024, 2025.
7. EMNLP Senior Area Chair 2024.
8. EMNLP Area Chair 2019, 2020, 2021, 2022, 2023.
9. ICLR Area Chair 2023, 2024.
10. ICML Area Chair 2023, 2024, 2025.
11. IJCAI Area Chair 2017, 2023.
12. KDD SPC 2022, 2023.
13. NAACL Senior Area Chair 2024, 2025.
14. NeurIPS Area Chair 2021, 2022, 2023, 2024, 2025.
15. NLPCC 2022 Area Chair on Machine Translation.
16. SDM SPC 2022.

Journal Reviewer: I was an active reviewer for JAIR, JMLR, and TACL.

Grant Reviewer: I was an reviewer for NSF SBIR panel in 2024 and European Research Council of European Union in 2022.

University Service

1. Undergraduate Curriculum committee of Computer Science Department, UCSB, 2022-2023.
2. Graduate Admission committee of Computer Science Department, UCSB, 2022-2023.
3. PR committee of Computer Science department, UCSB. 2021-2022.

Media Coverage

1. Open Source Neural Speech Translation Toolkit, Slator.com, Jan 7, 2021. link
2. Meet ByteDance AI's Xiaomingbot: World's First Multilingual and Multimodal AI News Agent, Synced Review, July21, 2020. link
3. Disrupted Journalism, by DisruptionHub. July 20, 2017. (mentioning our in-house developed Xiaomingbot) link
4. Robot gives competitive edge to tech application, by China Daily, September 4, 2016. (similar coverage by 83 reports. about our news writing robot: Xiaomingbot, some examples see below)
5. Prolific Robot Journalist Covers 450 Olympic Stories. by Newsweek. August 26, 2016. (coverage of our in-house developed writing robot Xiaomingbot) link
6. The Future of Writing? China's AI Reporter Published 450 Articles During Rio Olympics, by Futurism, August 26, 2016. link
7. Robot Reporter Covered The Olympics For A Chinese News Outlet, by Popular Science, August 25, 2016. (coverage of our in-house developed writing robot Xiaomingbot) link
8. China Sends AI Reporter to Cover the Rio Olympics – Blows Every Journalist Out of the Water, by Nextshark, August 25, 2016. link
9. A Chinese news outlet used an incredibly efficient "robot reporter" to cover the Olympics. by Quartz. August 24, 2016. link
10. China's A.I. Olympics Reporter Could Replace Journalists, by Inverse, August 24, 2016. link
11. Breaking down China's most popular news app, Toutiao, by AbacUSNews, Oct. 2, 2018. (mentioning Xiaomingbot) link
12. The Insanely Popular Chinese News App That You've Never Heard Of, by MIT Technology Review, January 26, 2017. (Mentioning Xiaomingbot) link
13. The Hidden Forces Behind Toutiao: China's Content King, by Y Combinator, October 12, 2017. link
14. Smart technologies hone poverty alleviation targeting, by Xinhua news, July 23, 2019. link (many similar coverage by various outlets)
15. Missing persons technology helps locate China's elderly, by ChinaDaily October 10, 2019. link (many similar coverage by various outlets)
16. Douyin took missing elder person home within three hours. by Sina.com, October 16, 2018. link (many similar coverage by various outlets)
17. Chinese new media urged to prioritize social effects, by ChinaDaily, November 30, 2019. link