Curriculum Vitae Last modified date: March 26, 2025

Kai-Wei Chang

University of California, Los Angeles kw@kwchang.net
ENG VI 374 http://kwchang.net
Los Angeles, CA 90095

RESEARCH INTERESTS

- Trustworthy NLP: We focus on aligning large language models (LLMs) with human values. Our lab has pioneered research on fairness and robustness in language representation and generation.
- Multimodal Foundation Models: We integrate vision and language representations. Our lab developed VisualBERT, one of the earliest vision-language foundation models, enabling AI to interpret and understand the visual world through language.
- Reasoning in NLP: We design methods that ensure LLMs adhere to specified constraints and enhance their commonsense, mathematical, and logical reasoning capabilities.

EDUCATION AND EXPERIENCE

University of California Los Angeles, CA
Associate Professor, Computer Science (current) 2022 –
Assistant Professor, Computer Science 2017 – 2022

Amazon.com, Inc.
Amazon Scholar (current) 2023 –
Amazon Visiting Academics 2020 – 2023

University of Virginia, VA
Assistant Professor, Computer Science 2016 – 2017

Microsoft, MA
Postdoctoral researcher, Microsoft Research New England Lab 2015 – 2016

University of Illinois at Urbana-Champaign, IL
Ph.D. in Computer Science (advisor: Dan Roth) 2010 – 2015

SELECTED AWARDS

AI's 10 to Watch, IEEE Intelligent Systems 2024

AAAI Senior Member 2023

Sloan Research Fellow 2021
the highest honors for an early career researcher.

Google Research Scholar Award 2021

NSF Research Initiation Initiative (CRII) Award 2016

EMNLP Best Long Paper Award (2017), KDD Best Paper Award (2010), ACL Outstanding Paper Award (2023), CVPR Best Paper Finalist (2022), Best paper awards at ACL-workshop (2017), ICLR-workshop(2023), ICLR-workshop (2024)

Okawa Research Grant Award 2018

FUNDING

I've secured funding support from DARPA, ONR, NSF, and several industry partners.

Multimodal InteRActive Conceptual Learning
$500k. DARPA ECOLE. Co-PI 2023-2026

NOVA: A Neuro-Symbolic Vision-Language Framework for Multimodal Human-Machine Interactions
$900k. ONR. PI 2023-2026

PYLON: An Integrated Semantic Framework for Probabilistic Neuro-Symbolic Learning and Reasoning DARPA ANSR. 2023-2027
$900k. Co-PI

Collaborative Research: SLES: Verifying and Enforcing Safety Constraints in AI-based Sequential Generation. 2023-2026
$270k. Co-PI

Optum Labs Research Grant
PI. $137k 2023
PI. $145k 2024
PI. $168k 2025

Taboola Research Gift Fund 2018–2026
$600k Solo-PI

PIPP Phase 1: An End-to-End Pandemic Early Warning System by Harnessing Open-Source Intelligence 2022-2024
NSF PIPP. Senior Personnel. PI: Wei Wang

Cisco Research Grant
Solo-PI. $75k (gift) 2024
Solo-PI. $165k 2023
Solo-PI. $153k 2022

Google Research Scholar
Solo-PI $60k (gift) 2021

Amazon Research Award 2020, 2023

AI-DCL: Governing Bias in AI System with Humans in the Decision Loop 2019–2021
NSF-IIS Eager grant. $300k. PI

Discovering Common Sense from Video, Images, Text and Knowledge Bases 2018–2019
DARPA MCS grant. $640k (my portion). PI of UCLA site. PI: Ralph Weischedel (USC)

Learning to Screen: Accelerating Training and Inference for Large NLP Models 2019–2020
Facebook Gift Grant. Co-PI with Cho-Jui Hsieh.

CICI: RDP: Security and Privacy Policy Enforcement for Research Data Protection 2019–2022
NSF-OAC grant. $210k (my portion). PI of UCLA site. PI: Yuan Tian (UVirginia)

Google GCP Credit Award 2019
$50,000 Google Cloud credits.

Discerning Group Biases in Online Communities via Linguistic Analysis 2018–2019
DARPA UGB grant. $300k (my portion). PI of UCLA site. PI: Aram Galstyan (USC)

Reducing Implicit Societal Bias in Artificial Intelligence Systems 2018
Research gift grant, The Okawa Foundataion. Solo-PI

CRII: RI: Learning Structured Prediction Models with Auxiliary Supervision 2016–2019
NSF-IIS grant. $170k. Solo-PI

PROFESSIONAL ACTIVITIES

Elected officer of ACL SIGDAT (the organizer of EMNLP). Vice President-Elect in 2024, Vice President in 2025, President in 2026.

Journal Action Editor: JAIR, JMLR, TACL, ACL Rolling Review

Conference Chairs:

- Demo Chair, NAACL 24
- Associate Program Chair, AAAI 23
- Ethics Committee Chair, NAACL 22
- Handbook Chair, EMNLP 18
- Senior Area Chair/Area Chair at ICML, NeurIPS, AAAI, NAACL, ACL, EMNLP, ICLR

Organizer:

- Southern California NLP symposium. UCLA, 2023 (156 posters, 300+ participants).
- TrustNLP: Workshop on Trustworthy Natural Language Processing. NAACL 21, 22, 24, 25, ACL 23.
- Multilingual Multimodal Learning (MML). ACL 2022.
- Robust and Reliable Machine Learning in the Real World. ICLR 2021.
- Workshop on Deep Structured Prediction. ICML 2017.
- Structured Prediction for Natural Language Processing. EMNLP 16, 17.

Tutorials

- Indirectly Supervised Natural Language Processing. ACL 2023.
- Fairness and Biases in Natural Language Processing. Machine Learning Summer School 2021.
- Robustness and Adversarial Examples in Natural Language Processing. EMNLP 2021.
- Recent Advances in Transferable Representation Learning, AAAI 2020.
- Bias and Fairness in Natural Language Processing, EMNLP 2019.
- Quantifying and Reducing Gender Stereotypes in Word Embeddings. FAccT 2018
- Structured Predictions: Practical Advancements and Applications in Natural Language Processing. TAAI 2017.
- Learning and Inference in Structured Prediction Models. AAAI 2016.
- Hands-on Learning to Search for Structured Prediction. NAACL 2015.

TEACHING EXPERIENCE

Instructor, CSM146: Introduction to Machine Learning, UCLA. Winter 18, Fall 19-22

Instructor, CS263: Natural Language Processing, UCLA. Spring 20-22

Instructor, CS269: Special Topic in AI: Fairness, Accountability, and Transparency in Natural Language Processing, UCLA. Winter 20-22

Instructor, CS269: Seminar: Machine Learning in Natural Language Processing, UCLA. Spring 2019, Fall 17.

Instructor, Advanced Machine Learning, University of Virginia. Fall 17.

ADVISEES

Postdoctoral Researcher

- James Wang

Ph.D. Students

- Amita Kamath, Graduate Dean Scholar Award and CS fellowship
- Cheng-Fu (Joey) Yang
- Christina Chance, Cota Robles Fellowship, GEM Fellowship
- Di Wu
- Fan Yin
- Da Yin, Amazon Fellowship
- Elaine Wan
- Zongyu Lin (Co-advised w/ Sun)
- Arjun Subramonian (Co-advised w/ Sun), Cota-Robles, NSF-Mentor Fellowships
- Tanmay Parekh (Co-advised w/ Peng), Amazon Fellowship (2024), Bloomberg Fellowship (2025)
- Hritik Bansal (Co-advised w/ Grover)
- Xueqing Wu (Co-advised w/ Peng)
- Yu (Bryan) Zhou
- Xingcheng Yao
- Rui Sun

Alumni in Academy

- Yining Hong (PhD, 2025), Postdoc at Stanford.
- Yiwei Wang (Postdoc, 2024) Current: Assistant Professor, UC-Merced
- Kareem Ahmad (PhD, 2024, Co-advised w/Broeck), Current: Postdoc at UCI
- Pan Lu (PhD, 2024, Co-advised w/ Zhu) Bloomberg, Amazon, Qualcomm Fellowships, Current: Postdoc at Stanford.
- Kuan-Hao Huang (PhD, 2023) Current: Assistant Professor at Texas A&M.
- Md. Rizwan Parvez (PhD, 2022) Current: Scientist at Qatar Computing Research Institute.
- Jieyu Zhao (PhD, 2021) Microsoft Fellowship, Current: Assistant Professor at USC.
- Ziniu Hu (PhD, 2022, co-advised w/Sun) Baidu, Amazon Fellowships, Current: Assistant Professor at RPI.
- Muhao Chen (PhD, 2018, unofficial advisee) Current: Assistant Professor at UCDavis.

Alumni in Industry and Research Lab

- Liunian Harold Li (PhD, 2024) Amazon, Google Fellowships, Current: Scientist at OpenAI.
- Tao Meng (PhD, 2024) Amazon Fellowship, Current: Scientist at Zoom.
- Anaelia Ovalle (PhD, 2024), Cota-Robles, NSF-Mentor Fellowships, Current: Meta
- Wasi Uddin Ahmad (PhD, 2021) Current: Scientist at Amazon AWS
- Sunipa Dev (Postdoc, 2022) Current: Scientist at Google.
- Dat Duong (PhD, 2020, co-advised w/ Eskin) Current: Scientist at NIH.

PUBLICATIONS

43,000+ Google Scholar citations. H-index: 73. Published at ACL, EMNLP, ICML, NeurIPS, AAAI...

Awarded, Spotlighted, and Highly Cited Papers[1]

[1] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, **Kai-Wei Chang**. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, **Best Long Paper Award, top 10 cited paper at EMNLP-17.**

[2] Hsiang-Fu Yu, Cho-Jui Hsieh, **Kai-Wei Chang**, and Chih-Jen Lin, Large linear classification when data cannot fit in memory, *the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2010),* **Best Research paper**.

[3] Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, Yiwu Zhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, **Kai-Wei Chang**, and Jianfeng Gao, Grounded Language-Image Pre-training, *the Computer Vision and Pattern Recognition Conference (CVPR 2022)*. **Best Paper Finalist, 33 out of 8161 submissions, top 0.4%**

[4] Kuan-Hao Huang, Varun Iyer, I.-Hung Hsu, Anoop Kumar, **Kai-Wei Chang**, and Aram Galstyan. "ParaAMR: A Large-Scale Syntactically Diverse Paraphrase Dataset by AMR Back-Translation." In the Annual Meeting of the Association for Computational Linguistics (ACL 2023). **Area Chair Award (top paper in Semantics track)**

[5] Nikil Roashan Selvam, Sunipa Dev, Daniel Khashabi, Tushar Khot, and **Kai-Wei Chang**. "The Tail Wagging the Dog: Dataset Construction Biases of Social Bias Benchmarks." In the Annual Meeting of the Association for Computational Linguistics (ACL short), 2023. **Outstanding paper Award**

[6] Hritik Bansal, Yonatan Bitton, Idan Szpektor, **Kai-Wei Chang**, and Aditya Grover. *VideoCon: Robust Video-Language Alignment via Contrast Captions*. In the Conference on Computer Vision and Pattern Recognition (CVPR, 2024). **Best Paper Award at ICLR Workshop on Data Problems for Foundation Models.**

[7] Pan Lu, Hritik Bansal, Tony Xia, Jiacheng Liu, Chunyuan Li, Hannaneh Hajishirzi, Hao Cheng, **Kai-Wei Chang**, Michel Galley, and Jianfeng Gao. *MathVista: Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts*. In the International Conference on Learning Representations (ICLR, 2024).**Selected for oral (85 out of 7,000 submissions, top 1.2%)**

[8] Hritik Bansal, Nishad Singhi, Yu Yang, Fan Yin, Aditya Grover, and **Kai-Wei Chang**. "CleanCLIP: Mitigating Data Poisoning Attacks in Multimodal Contrastive Learning." In the International Conference on Computer Vision (ICCV, 2023). **Selected for oral (195 out of 8088 submissions, top 2.5%), Best Paper Award at ICLR Workshop on Trustworthy and Reliable Large-Scale Machine Learning Models.**

[9] Tao Meng, Sidi Lu, Nanyun Peng, and **Kai-Wei Chang**. Controllable Text Generation with Neurally-Decomposed Oracle In *Neural Information Processing Systems (NeurIPS 2022).* **Selected for oral, 201 out of 10411, top 1.9%**

[10] Shaochen Zhong, Yifan Lu, Lize Shao, Bhargav Bhushanam, Xiaocong Du, Yixin Wan, Yucheng Shi, Daochen Zha, Yiwei Wang, Ninghao Liu, Kaixiong Zhou, Shuai Xu, **Kai-Wei Chang**, Louis Feng, Vipin Chaudhary, and others. âĂIJMQuAKE-Remastered: Multi-Hop Knowledge Editing Can Only Be Advanced with Reliable Evaluations.âĂİ In ICLR, 2025. **Selected for spotlight, 373 out of 11672, top 5%**

[1] Measured by citation counts up to May 2024 according to Paper Digest https://www.paperdigest.org/best-paper-digest/

[11] Yining Hong, Beide Liu, Maxine Wu, Yuanhao Zhai, **Kai-Wei Chang**, Linjie Li, Kevin Lin, Chung-Ching Lin, Jianfeng Wang, Zhengyuan Yang, Ying Nian Wu, and Lijuan Wang. âĂIJSlowFast-VGen: Slow-Fast Learning for Action-Driven Long Video Generation.âĂİ In ICLR, 2025. **Selected for spotlight, 373 out of 11672, top 5%**

[12] Honghua Zhang, Liunian Harold Li, Tao Meng, **Kai-Wei Chang**, and Guy Van den Broeck. "On the Paradox of Learning to Reason from Data." In *the International Joint Conferences on Artificial Intelligence (IJCAI 2023)*. **top-3 cited paper in IJCAI-23**

[13] Sheng Shen, Liunian Harold Li, Hao Tan, Mohit Bansal, Anna Rohrbach, **Kai-Wei Chang**, Zhewei Yao, and Kurt Keutz, How Much Can CLIP Benefit Vision-and-Language Tasks? *International Conference on Learning Representation (ICLR 2022).* **top 10 cited paper at ICLR-22**

[14] Pan Lu, Swaroop Mishra, Tony Xia, Liang Qiu, **Kai-Wei Chang**, Song-Chun Zhu, Oyvind Tafjord, Peter Clark, and Ashwin Kalyan. Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering. In *Neural Information Processing Systems (NeurIPS 2022).* **top 15 cited paper at NeurIPS-22**

[15] W. Ahmad, S. Chakraborty, B. Ray, **K.-W. Chang**. Unified Pre-training for Program Understanding and Generation. *North American Chapter of the Association for Computational Linguistics (NAACL 2021)*, **top 3 cited paper at NAACL-21**.

[16] Z. Hu, Y. Dong, K. Wang, **K.-W. Chang**, and Y. Sun. GPT-GNN: Generative Pre-Training of Graph Neural Networks. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2020)*, **top 10 cited paper at KDD-20**.

[17] M. Alzantot, Y. Sharma, A. Elgohary, B.-J. Ho, M. Srivastava, **K.-W. Chang**. Generating Natural Language Adversarial Examples. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2018, short)*, **top 5 cited paper at EMNLP-18**

[18] J. Zhao, T. Wang, M. Yatskar, V. Ordonez, **K.-W. Chang**, Gender Bias in Coreference Resolution:Evaluation and Debiasing Methods. *North American Chapter of the Association for Computational Linguistics (NAACL 2018, short)*, **top 10 cited paper at NAACL-18** .

[19] T. Bolukbasi, **K.-W Chang** , James Zou, Venkatesh Saligrama, Adam Kalai, Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Neural Information Processing Systems (NeurIPS 2016).* **top 10 cited paper at NeurIPS-16.**

[20] C.-J. Hsieh, **K.-W. Chang**, C.-J. Lin, S. Sathiya Keerthi, and S. Sundararajan, A dual coordinate descent method for large-scale linear SVM, *the 25th International Conference on Machine Learning (ICML 2008).* **top 5 cited paper at ICML-08.**

Other Refereed Conference Publications

[21] Xueqing Wu, Yuheng Ding, Bingxuan Li, Pan Lu, Da Yin, **Kai-Wei Chang**, and Nanyun Peng. "VISCO: Benchmarking Fine-Grained Critique and Correction Towards Self-Improvement in Visual Reasoning." In CVPR, 2025.

[22] Shuyang Hao, Bryan Hooi, Jun Liu, **Kai-Wei Chang**, Zi Huang, and Yujun Cai. "Exploring Visual Vulnerabilities via Multi-Loss Adversarial Search for Jailbreaking Vision-Language Models." In CVPR, 2025.

[23] Hritik Bansal, Zongyu Lin, Tianyi Xie, Zeshun Zong, Michal Yarom, Yonatan Bitton, Chenfanfu Jiang, Yizhou Sun, **Kai-Wei Chang**, and Aditya Grover. "VideoPhy: Evaluating Physical Commonsense for Video Generation." In ICLR, 2025.

[24] Fei Wang, Xingyu Fu, James Y. Huang, Zekun Li, Qin Liu, Xiaogeng Liu, Mingyu Derek Ma, Nan Xu, Wenxuan Zhou, Kai Zhang, Tianyi Lorena Yan, Wenjie Jacky Mo, Hsiang-Hui Liu, Pan Lu,

Chunyuan Li, and others. "MuirBench: A Comprehensive Benchmark for Robust Multi-image Understanding." In ICLR, 2025.

[25] Kareem Ahmed, **Kai-Wei Chang**, and Guy Van den Broeck. "Controllable Generation via Locally Constrained Resampling." In ICLR, 2025.

[26] Wenbo Hu, Jia-Chen Gu, Zi-Yi Dou, Mohsen Fayyaz, Pan Lu, **Kai-Wei Chang**, and Nanyun Peng. "MRAG-Bench: Vision-Centric Evaluation for Retrieval-Augmented Multimodal Models." In ICLR, 2025.

[27] Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, **Kai-Wei Chang**, and Dong Yu. "LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory." In ICLR, 2025.

[28] Xiaomeng Jin, Zhiqi Bu, Bhanukiran Vinzamuri, Anil Ramakrishna, **Kai-Wei Chang**, Volkan Cevher, and Mingyi Hong. "Unlearning as Multi-task Optimization: A Normalized Gradient Difference Approach with an Adaptive Learning Rate." In NAACL, 2025.

[29] Anubrata Das, Manoj Kumar, Ninareh Mehrabi, Anil Ramakrishna, Anna Rumshisky, **Kai-Wei Chang**, Aram Galstyan, Morteza Ziyadi, and Rahul Gupta. "On Localizing and Deleting Toxic Memories in Large Language Models." In NAACL-Finding, 2025.

[30] Yiwei Wang, Muhao Chen, Nanyun Peng, and **Kai-Wei Chang**. "Vulnerability of Large Language Models to Output Prefix Jailbreaks: Impact of Positions on Safety." In NAACL-Finding, 2025.

[31] Yuankai Li, Jia-Chen Gu, Di Wu, **Kai-Wei Chang**, and Nanyun Peng. "BRIEF: Bridging Retrieval and Inference for Multi-hop Reasoning via Compression." In NAACL-Finding, 2025.

[32] Yihe Deng, Pan Lu, Fan Yin, Ziniu Hu, Sheng Shen, Quanquan Gu, James Zou, **Kai-Wei Chang**, and Wei Wang. "Enhancing Large Vision Language Models with Self-Training on Image Comprehension." In NeurIPS, 2024.

[33] Wenbo Hu, Zi-Yi Dou, Liunian Harold Li, Amita Kamath, Nanyun Peng, and **Kai-Wei Chang**. "MQT-LLaVA: Matryoshka Query Transformer for Large Vision-Language Models." In NeurIPS, 2024.

[34] Da Yin, Haoyi Qiu, Kung-Hsiang Huang, **Kai-Wei Chang**, and Nanyun Peng. "SafeWorld: Geo-Diverse Safety Alignment." In NeurIPS, 2024.

[35] Zhecan Wang, Junzhang Liu, Chia-Wei Tang, Hani Alomari, Anushka Sivakumar, Rui Sun, Wenhao Li, Md. Atabuzzaman, Hammad Ayyubi, Haoxuan You, Alvi Md Ishmam, **Kai-Wei Chang**, Shih-Fu Chang, and Chris Thomas. "JourneyBench: A Challenging One-Stop Vision-Language Understanding Benchmark of Generated Images." In NeurIPS (Datasets and Benchmarks Track), 2024.

[36] Xueqing Wu, Rui Zheng, Jingzhen Sha, Te-Lin Wu, Hanyu Zhou, Tang Mohan, **Kai-Wei Chang**, Nanyun Peng, and Haoran Huang. "DACO: Towards Application-Driven and Comprehensive Data Analysis via Code Generation." In NeurIPS (Datasets and Benchmarks Track), 2024.

[37] Di Wu, Jia-Chen Gu, Fan Yin, Nanyun Peng, and **Kai-Wei Chang**. "Synchronous Faithfulness Monitoring for Trustworthy Retrieval-Augmented Generation." In EMNLP, 2024.

[38] Ninareh Mehrabi, Palash Goyal, Christophe Dupuy, Qian Hu, Shalini Ghosh, Richard Zemel, **Kai-Wei Chang**, Aram Galstyan, and Rahul Gupta. "FLIRT: Feedback Loop In-context Red Teaming." In EMNLP, 2024.

[39] Ashima Suvarna, Xiao Liu, Tanmay Parekh, **Kai-Wei Chang**, and Nanyun Peng. "QUDSELECT: Selective Decoding for Questions Under Discussion Parsing." In EMNLP, 2024.

[40] Fei Wang, Ninareh Mehrabi, Palash Goyal, Rahul Gupta, **Kai-Wei Chang**, and Aram Galstyan. "Data Advisor: Data Curation with Foresight for Safety Alignment of Large Language Models." In EMNLP, 2024.

[41] Yixin Wan, Di Wu, Haoran Wang, and **Kai-Wei Chang**. "The Factuality Tax of Diversity-Intervened Text-to-Image Generation: Benchmark and Fact-Augmented Intervention." In EMNLP, 2024.

[42] Tanmay Parekh, Jeffrey Kwan, Jiarui Yu, Sparsh Johri, Hyosang Ahn, Sreya Muppalla, **Kai-Wei Chang**, Wei Wang, and Nanyun Peng. "SPEED++: A Multilingual Event Extraction Framework for Epidemic Prediction and Preparedness." In EMNLP, 2024.

[43] Bingxuan Li, Yiwei Wang, Tao Meng, Nanyun Peng, and **Kai-Wei Chang**. "Evaluating LLMsâĂŹ Capability in Satisfying Lexical Constraints." In EMNLP, 2024.

[44] Zi-Yi Dou, Cheng-Fu Yang, Xueqing Wu, **Kai-Wei Chang**, and Nanyun Peng. "Re-ReST: Reflection-Reinforced Self-Training for Language Agents." In EMNLP, 2024.

[45] Jia-Chen Gu, Hao-Xiang Xu, Jun-Yu Ma, Pan Lu, Zhen-Hua Ling, **Kai-Wei Chang**, and Nanyun Peng. "Model Editing Harms General Abilities of Large Language Models: Regularization to the Rescue." In EMNLP, 2024.

[46] Di Wu, Xiaoxian Shen, and **Kai-Wei Chang**. "MetaKP: On-Demand Keyphrase Generation." In EMNLP-Findings, 2024.

[47] Tao Meng, Ninareh Mehrabi, Palash Goyal, Anil Ramakrishna, Aram Galstyan, Richard Zemel, **Kai-Wei Chang**, Rahul Gupta, and Charith Peris. "Attribute Controlled Fine-tuning for Large Language Models: A Case Study on Detoxification." In EMNLP-Findings, 2024.

[48] Silin Meng, Yiwei Wang, Cheng-Fu Yang, Nanyun Peng, and **Kai-Wei Chang**. "LLM-A*: Large Language Model Enhanced Incremental Heuristic Search on Path Planning." In EMNLP-Findings, 2024.

[49] Zongyu Lin, Zhikun Xu, Yixin Wan, Stuart X. Yao, Xiaohan Song, Tsung-Han Lin, Selina Song, Pranav Subbaraman, **Kai-Wei Chang**, and Yizhou Sun. "VISUAL-ALPHASOCIAL: Benchmark and Self-Reflective Chain-of-Thought Generation for Visual Social Commonsense Reasoning." In EMNLP-Findings, 2024.

[50] Xueqing Wu, Zongyu Lin, Songyan Zhao, Te-Lin Wu, Pan Lu, Nanyun Peng, and **Kai-Wei Chang**. "VDebugger: Harnessing Execution Feedback for Debugging Visual Programs." In EMNLP-Findings, 2024.

[51] Shujin Wu, Yi Fung, Sha Li, Yixin Wan, **Kai-Wei Chang**, and Heng Ji. "MACAROON: Training Vision-Language Models To Be Your Engaged Partners." In EMNLP-Findings, 2024.

[52] Renrui Zhang, Dongzhi Jiang, Yichi Zhang, Haokun Lin, Pengshuo Qiu, Ziyu Guo, Aojun Zhou, Pan Lu, **Kai-Wei Chang**, Peng Gao, and Hongsheng Li. "MathVerse: Does Your Multi-modal LLM Truly See the Diagrams in Visual Math Problems?" In the European Conference on Computer Vision (ECCV 2024).

[53] Amita Kamath, Cheng-Yu Hsieh, **Kai-Wei Chang**, and Ranjay Krishna. "The Hard Positive Truth about Vision-Language Compositionality." In the European Conference on Computer Vision (ECCV 2024).

[54] Di Wu, Da Yin, **Kai-Wei Chang**. "KPEval: Towards Fine-Grained Semantic-Based Keyphrase Evaluation." In the Annual Meeting of the Association for Computational Linguistics (ACL Findings 2024).

[55] Xiao Liu, Zirui Wu, Xueqing Wu, Pan Lu, **Kai-Wei Chang**, and Yansong Feng. "Are LLMs Capable of Data-based Statistical and Causal Reasoning? Benchmarking Advanced Quantitative Reasoning with Data." In the Annual Meeting of the Association for Computational Linguistics (ACL Findings 2024).

[56] Elan Sopher Markowitz, Anil Ramakrishna, Jwala Dhamala, Ninareh Mehrabi, Charith Peris, Rahul Gupta, **Kai-Wei Chang**, and Aram Galstyan. "Tree-of-Traversals: A Zero-Shot Reasoning

Algorithm for Augmenting Black-box Language Models with Knowledge Graphs." In the Annual Meeting of the Association for Computational Linguistics (ACL 2024).

[57] Da Yin, Faeze Brahman, Abhilasha Ravichander, Khyathi Chandu, **Kai-Wei Chang**, Yejin Choi, and Bill Yuchen Lin. "Agent Lumos: Unified and Modular Training for Open-Source Language Agents." In the Annual Meeting of the Association for Computational Linguistics (ACL 2024).

[58] Kuan-Hao Huang, I-Hung Hsu, Tanmay Parekh, Zhiyu Xie, Zixuan Zhang, Prem Natarajan, **Kai-Wei Chang**, Nanyun Peng, and Heng Ji. "TextEE: Benchmark, Reevaluation, Reflections, and Future Challenges in Event Extraction." In the Annual Meeting of the Association for Computational Linguistics (ACL Findings 2024).

[59] Fan Yin, Jayanth Srinivasa, and **Kai-Wei Chang**. *Characterizing Truthfulness in Large Language Model Generations with Local Intrinsic Dimension*. In the International Conference on Machine Learning (ICML, 2024).

[60] Rohan Wadhawan, Hritik Bansal, **Kai-Wei Chang**, and Nanyun Peng. *ConTextual: Evaluating Context-Sensitive Text-Rich Visual Reasoning in Large Multimodal Models*. In the International Conference on Machine Learning (ICML, 2024).

[61] Chujie Zheng, Fan Yin, Hao Zhou, Fandong Meng, Jie Zhou, **Kai-Wei Chang**, Minlie Huang, and Nanyun Peng. *Prompt-Driven LLM Safeguarding via Directed Representation Optimization*. In the International Conference on Machine Learning (ICML, 2024).

[62] Yue Huang, Lichao Sun, Haoran Wang, Siyuan Wu, Qihui Zhang, Yuan Li, Chujie Gao, Yixin Huang, Wenhan Lyu, Yixuan Zhang, Xiner Li, Hanchi Sun, Zhengliang Liu, Yixin Liu, Yijue Wang, Zhikun Zhang, Bertie Vidgen, Bhavya Kailkhura, Caiming Xiong, Chaowei Xiao, Chunyuan Li, Eric P. Xing, Furong Huang, Hao Liu, Heng Ji, Hongyi Wang, Huan Zhang, Huaxiu Yao, Manolis Kellis, Marinka Zitnik, Meng Jiang, Mohit Bansal, James Zou, Jian Pei, Jian Liu, Jianfeng Gao, Jiawei Han, Jieyu Zhao, Jiliang Tang, Jindong Wang, Joaquin Vanschoren, John Mitchell, Kai Shu, Kaidi Xu, **Kai-Wei Chang**, Lifang He, Lifu Huang, Michael Backes, Neil Zhenqiang Gong, Philip S. Yu, Pin-Yu Chen, Quanquan Gu, Ran Xu, Rex Ying, Shuiwang Ji, Suman Jana, Tianlong Chen, Tianming Liu, Tianyi Zhou, William Yang Wang, Xiang Li, Xiangliang Zhang, Xiao Wang, Xing Xie, Xun Chen, Xuyu Wang, Yan Liu, Yanfang Ye, Yinzhi Cao, Yong Chen, and Yue Zhao. *TrustLLM: Trustworthiness in Large Language Models*. In the International Conference on Machine Learning (ICML, 2024).

[63] Mingyu Derek Ma, Jiun-Yu Kao, Arpit Gupta, Yu-Hsiang Lin, Wenbo Zhao, Tagyoung Chung, Wei Wang, **Kai-Wei Chang**, and Nanyun Peng. *Mitigating Bias for Question Answering Models by Tracking Bias Influence*. In the North American Chapter of the Association for Computational Linguistics (NAACL, 2024).

[64] Xiao Liu, Yansong Feng, and **Kai-Wei Chang**. *CASA: Causality-driven Argument Sufficiency Assessment*. In the North American Chapter of the Association for Computational Linguistics (NAACL, 2024).

[65] Tanmay Parekh, I.-Hung Hsu, Kuan-Hao Huang, **Kai-Wei Chang**, and Nanyun Peng. *Contextual Label Projection for Cross-Lingual Structured Prediction*. In the North American Chapter of the Association for Computational Linguistics (NAACL, 2024).

[66] Tanmay Parekh, Anh Mac, Jiarui Yu, Yuxuan Dong, Syed Shahriar, Bonnie Liu, Eric J. Yang, Kuan-Hao Huang, Wei Wang, Nanyun Peng, and **Kai-Wei Chang**. *Event Detection from Social Media for Epidemic Prediction*. In the North American Chapter of the Association for Computational Linguistics (NAACL, 2024).

[67] Junyi Li, Ninareh Mehrabi, Charith Peris, Palash Goyal, **Kai-Wei Chang**, Aram Galstyan, Richard Zemel, and Rahul Gupta. *The Steerability of Large Language Models Toward Data-Driven*

Personas. In the North American Chapter of the Association for Computational Linguistics (NAACL, 2024).

[68] Di Wu, Wasi Uddin Ahmad, and **Kai-Wei Chang**. *On Leveraging Encoder-Only Pre-trained Language Models for Effective Keyphrase Generation*. In the International Conference on Language Resources and Evaluation Conference (LREC-COLING, 2024).

[69] Anaelia Ovalle, Ninareh Mehrabi, Palash Goyal, Jwala Dhamala, **Kai-Wei Chang**, Richard Zemel, Aram Galstyan, Yuval Pinter, and Rahul Gupta. *Are You Talking to ['xem'] or ['x', 'em']? On Tokenization and Addressing Misgendering in LLMs with Pronoun Tokenization Parity*. In the North American Chapter of the Association for Computational Linguistics - Findings (NAACL-Findings, 2024).

[70] Jooyoung Lee, Fan Yang, Thanh Tran, Qian Hu, Emre Barut, **Kai-Wei Chang**, and Chengwei Su. *Can Small Language Models Help Large Language Models Reason Better?: LM-Guided Chain-of-Thought*. In the International Conference on Language Resources and Evaluation Conference (LREC-COLING, 2024).

[71] Haoxuan You, Mandy Guo, Zhecan Wang, **Kai-Wei Chang**, Jason Michael Baldridge, and Jiahui Yu. *CoBIT: A Contrastive Bi-Directional Image-Text Generation Model*. In the International Conference on Learning Representations (ICLR, 2024).

[72] Oscar Chew, Hsuan-Tien Lin, **Kai-Wei Chang**, and Kuan-Hao Huang. *Understanding and Mitigating Spurious Correlations in Text Classification with Neighborhood Analysis*. In the European Chapter of the Association for Computational Linguistics - Findings (EACL-Findings, 2024).

[73] Po-Nien Kung, Fan Yin, Di Wu, **Kai-Wei Chang**, and Nanyun Peng. "Active Instruction Tuning: Improving Cross-Task Generalization by Training on Prompt Sensitive Tasks." In the Conference on Empirical Methods in Natural Language Processing (EMNLP, 2023).

[74] Amita Kamath, Jack Hessel, and **Kai-Wei Chang**. "What's 'up' with vision-language models? Investigating their struggle to understand spatial relations." In the Conference on Empirical Methods in Natural Language Processing (EMNLP, 2023).

[75] Amita Kamath, Jack Hessel, and **Kai-Wei Chang**. "Text Encoders are Performance Bottlenecks in Contrastive Vision-Language Models." In the Conference on Empirical Methods in Natural Language Processing (EMNLP, 2023).

[76] Di Wu, Wasi Uddin Ahmad, and **Kai-Wei Chang**. "Rethinking Model Selection and Decoding for Keyphrase Generation with Pre-trained Sequence-to-Sequence Models." In the Conference on Empirical Methods in Natural Language Processing (EMNLP, 2023).

[77] Cheng-Fu Yang, Yen-Chun Chen, Jianwei Yang, Xiyang Dai, Lu Yuan, Yu-Chiang Frank Wang, and **Kai-Wei Chang**. "LACMA: Language-Aligning Contrastive Learning with Meta-Actions for Embodied Instruction Following." In the Conference on Empirical Methods in Natural Language Processing (EMNLP, 2023).

[78] Da Yin, Xiao Liu, Fan Yin, Ming Zhong, Hritik Bansal, Jiawei Han, and **Kai-Wei Chang**. "Dynosaur: A Dynamic Growth Paradigm for Instruction-Tuning Data Curation." In the Conference on Empirical Methods in Natural Language Processing (EMNLP, 2023).

[79] Haoxuan You, Rui Sun, Zhecan Wang, Long Chen, Gengyu Wang, Hammad Ayyubi, **Kai-Wei Chang**, and Shih-Fu Chang. "IdealGPT: Iteratively Decomposing Vision and Language Reasoning via Large Language Models." In the EMNLP-Finding, 2023.

[80] Yixin Wan, Jieyu Zhao, Aman Chadha, Nanyun Peng, and **Kai-Wei Chang**. "Are Personalized Stochastic Parrots More Dangerous? Evaluating Persona Biases in Dialogue Systems." In the EMNLP-Finding, 2023.

[81] Yixin Wan, George Pu, Jiao Sun, Aparna Garimella, **Kai-Wei Chang**, and Nanyun Peng. "Kelly is a Warm Person, Joseph is a Role Model: Gender Biases in LLM-Generated Reference Letters." In the EMNLP-Findings, 2023.

[82] Zhecan Wang, Long Chen, Haoxuan You, Keyang Xu, Noel C. Codella, **Kai-Wei Chang**, and Shih-Fu Chang. "Dataset Bias Mitigation in Multiple-Choice Visual Question Answering and Beyond." In the EMNLP-Findings, 2023.

[83] Liunian Harold Li, Zi-Yi Dou, Nanyun Peng, and **Kai-Wei Chang**. "DesCo: Learning Object Recognition with Rich Language Descriptions." In the Neural Information Processing Systems (NeurIPS, 2023).

[84] Ziniu Hu, Ahmet Iscen, Chen Sun, **Kai-Wei Chang**, Yizhou Sun, Cordelia Schmid, David A. Ross, and Alireza Fathi. "AVIS: Autonomous Visual Information Seeking with Large Language Models." In the Neural Information Processing Systems (NeurIPS, 2023).

[85] Pan Lu, Baolin Peng, Hao Cheng, Michel Galley, **Kai-Wei Chang**, Ying Nian Wu, Song-Chun Zhu, and Jianfeng Gao. "Chameleon: Plug-and-Play Compositional Reasoning with Large Language Models." In the Neural Information Processing Systems (NeurIPS, 2023).

[86] Kareem Ahmed, **Kai-Wei Chang**, and Guy Van den Broeck. "A Pseudo-Semantic Loss for Deep Generative Models with Logical Constraints." In the Neural Information Processing Systems (NeurIPS, 2023).

[87] Liunian Harold Li, Jack Hessel, Youngjae Yu, Xiang Ren, **Kai-Wei Chang**, and Yejin Choi. "Symbolic Chain-of-Thought Distillation: Small Models Can Also 'Think' Step-by-Step." In the Annual Meeting of the Association for Computational Linguistics (ACL 2023).

[88] Tanmay Parekh, I.-Hung Hsu, Kuan-Hao Huang, **Kai-Wei Chang**, and Nanyun Peng. "GENEVA: Pushing the Limit of Generalizability for Event Argument Extraction with 100+ Event Types." In the Annual Meeting of the Association for Computational Linguistics (ACL 2023).

[89] Chenghao Yang, Fan Yin, He He, **Kai-Wei Chang**, Xiaofei Ma, and Bing Xiang. "Efficient Shapley Values Estimation by Amortization for Text Classification." In the Annual Meeting of the Association for Computational Linguistics (ACL 2023).

[90] I.-Hung Hsu, Kuan-Hao Huang, Shuning Zhang, Wenxin Cheng, Prem Natarajan, **Kai-Wei Chang**, and Nanyun Peng. "TAGPRIME: A Unified Framework for Relational Structure Extraction." In the Annual Meeting of the Association for Computational Linguistics (ACL 2023).

[91] Ninareh Mehrabi, Palash Goyal, Apurv Verma, Jwala Dhamala, Varun Kumar, Qian Hu, **Kai-Wei Chang**, Richard Zemel, Aram Galstyan, and Rahul Gupta. "Resolving Ambiguities in Text-to-Image Generative Models." In the Annual Meeting of the Association for Computational Linguistics (ACL 2023).

[92] Pan Lu, Liang Qiu, Wenhao Yu, Sean Welleck, and **Kai-Wei Chang**. "A Survey of Deep Learning for Mathematical Reasoning." In the Annual Meeting of the Association for Computational Linguistics (ACL 2023).

[93] Masoud Monajatipoor, Liunian Harold Li, Mozhdeh Rouhsedaghat, Lin Yang, and **Kai-Wei Chang**. "Transferring In-Context Learning Ability From Language Models to Vision-Language Models." In the Annual Meeting of the Association for Computational Linguistics (ACL short), 2023.

[94] Jianfeng Chi, Wasi Uddin Ahmad, Yuan Tian, and **Kai-Wei Chang**. "PLUE: Language Understanding Evaluation Benchmark for Privacy Policies in English." In the Annual Meeting of the Association for Computational Linguistics (ACL short), 2023.

[95] Nikil Roashan Selvam, Sunipa Dev, Daniel Khashabi, Tushar Khot, and **Kai-Wei Chang**. "The

Tail Wagging the Dog: Dataset Construction Biases of Social Bias Benchmarks." In the Annual Meeting of the Association for Computational Linguistics (ACL short), 2023.

[96] Yixin Wan, Kuan-Hao Huang, and **Kai-Wei Chang**. "PIP: Parse-Instructed Prefix for Syntactically Controlled Paraphrase Generation." In the Annual Meeting of the Association for Computational Linguistics (ACL short), 2023.

[97] Zixuan Ling, Xiaoqing Zheng, Jianhan Xu, Jinshu Lin, **Kai-Wei Chang**, Cho-Jui Hsieh, and Xuanjing Huang. "Enhancing Unsupervised Semantic Parsing with Distributed Contextual Representations." In *the Annual Meeting of the Association for Computational Linguistics (ACL-Finding 2023)*.

[98] Rui Sun, Zhecan Wang, Haoxuan You, Noel Codella, **Kai-Wei Chang**, and Shih-Fu Chang. "UniFine: A Unified and Fine-grained Approach for Zero-shot Vision-Language Understanding." In *the Annual Meeting of the Association for Computational Linguistics (ACL-Finding 2023)*.

[99] Wasi Ahmad, Md Golam Rahman Tushar, Saikat Chakraborty, and **Kai-Wei Chang**. "AVATAR: A Parallel Corpus for Java-Python Program Translation." In *the Annual Meeting of the Association for Computational Linguistics (ACL-Finding, short, 2023)*.

[100] Ziniu Hu, Ahmet Iscen, Chen Sun, Zirui Wang, **Kai-Wei Chang**, Yizhou Sun, Cordelia Schmid, David A. Ross, and Alireza Fathi. "REVEAL: Retrieval-Augmented Visual-Language Pre-Training with Multi-Source Multimodal Knowledge." In *the Computer Vision and Pattern Recognition Conference (CVPR 2023)*.

[101] Da Yin, Feng Gao, Govind Thattai, Michael Johnston, and **Kai-Wei Chang**. "GIVL: On Improving Geographical Inclusivity of Vision-and-Language Models with Pre-Training Methods." In *the Computer Vision and Pattern Recognition Conference (CVPR 2023)*.

[102] Kareem Ahmed, **Kai-Wei Chang**, and Guy Van den Broeck. "Semantic Strengthening of Neuro-Symbolic Learning." *In the International Conference on Artificial Intelligence and Statistics (AISTATS 2023)*.

[103] Anaelia Ovalle, Arjun Subramonian, Vagrant Gautam, Gilbert Gee, and **Kai-Wei Chang**. "Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness." In *AAAI/ACM Conference on AI, Ethics, and Society (AIES 2023)*.

[104] Pan Lu, Liang Qiu, **Kai-Wei Chang**, Ying Nian Wu, Song-Chun Zhu, Tanmay Rajpurohit, Peter Clark, and Ashwin Kalyan. Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning. *International Conference on Learning Representation (ICLR 2023).*

[105] Fan Yin, Yao Li, Cho-Jui Hsieh, and **Kai-Wei Chang**. ADDMU: Detection of Far-Boundary Adversarial Examples with Data and Model Uncertainty Estimation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*.

[106] Zhecan Wang, Haoxuan You, Yicheng He, Wenhao Li,**Kai-Wei Chang**, and Shih-Fu Chang. Understanding ME? Multimodal Evaluation for Fine-grained Visual Commonsense In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*.

[107] Ziniu Hu, Yichong Xu, Wenhao Yu, Shuohang Wang, Ziyi Yang, Chenguang Zhu, **Kai-Wei Chang**, and Yizhou Sun. Empowering Language Models with Knowledge Graph Reasoning for Open-Domain Question Answering. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*.

[108] Da Yin, Hritik Bansal, Masoud Monajatipoor, Liunian Harold Li, and **Kai-Wei Chang**. GeoMLAMA: Geo-Diverse Commonsense Probing on Multilingual Pre-Trained Language Models. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*.

[109] Hritik Bansal, Da Yin, Masoud Monajatipoor, and **Kai-Wei Chang**. How well can Text-to-Image Generative Models understand Ethical Natural Language Interventions?

[110] Ziniu Hu, Yichong Xu, Wenhao Yu, Shuohang Wang, Ziyi Yang, Chenguang Zhu, **Kai-Wei Chang**, and Yizhou Sun. Empowering Language Models with Knowledge Graph Reasoning for Open-Domain Question Answering. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022, short)*.

[111] Haoxuan You, Rui Sun, Zhecan Wang, **Kai-Wei Chang**, and Shih-Fu Chang. Find Someone Who: Visual Commonsense Understanding in Human-Centric Grounding. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022, Finding)*.

[112] Jieyu Zhao, Xuezhi Wang, Yao Qin, Jilin Chen, and **Kai-Wei Chang**. Investigating Ensemble Methods for Model Robustness Improvement of Text Classifiers. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022, Finding short)*.

[113] Jianfeng Chi, William Shand, Yaodong Yu, **Kai-Wei Chang**, Han Zhao, and Yuan Tian. Conditional Supervised Contrastive Learning for Fair Text Classification. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022, finding)*.

[114] Di Wu, Wasi Uddin Ahmad, Sunipa Dev, and **Kai-Wei Chang**. Representation Learning for Resource-Constrained Keyphrase Generation In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2022, finding)*.

[115] Jwala Dhamala, Varun Kumar, Rahul Gupta, **Kai-Wei Chang**, Aram Galstyan. An analysis of the effects of decoding algorithms on fairness in open-ended language generation. In IEEE Spoken Language Technology Workshop (SLT 2022)

[116] Arjun Subramonian, **Kai-Wei Chang**, and Yizhou Sun. On the Discrimination Risk of Mean Aggregation Feature Imputation in Graphs In *Neural Information Processing Systems (NeurIPS 2022).*

[117] Kareem Ahmed, Stefano Teso, **Kai-Wei Chang**, Guy Van den Broeck, and Antonio Vergari. Semantic Probabilistic Layers for Neuro-Symbolic Learning. In *Neural Information Processing Systems (NeurIPS 2022).*

[118] Kareem Ahmed, Eric Wang, **Kai-Wei Chang**, and Guy Van den Broeck, Neuro-Symbolic Entropy Regularization. *Uncertainty in Artificial Intelligence (UAI 2022).*

[119] Vijit Malik, Sunipa Dev, Akihiro Nishi, Nanyun Peng, and **Kai-Wei Chang**, Socially Aware Bias Measurements for Hindi Language Representations. *North American Chapter of the Association for Computational Linguistics (NAACL short 2022)*.

[120] I.-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Prem Natarajan, **Kai-Wei Chang**, and Nanyun Peng. DEGREE: A Data-Efficient Generative Event Extraction Model. *North American Chapter of the Association for Computational Linguistics (NAACL 2022)*.

[121] Kuan-Hao Huang, I.-Hung Hsu, Prem Natarajan, **Kai-Wei Chang**, and Nanyun Peng, Multilingual Generative Language Models for Zero-Shot Cross-Lingual Event Argument Extraction *the Annual Meeting of the Association for Computational Linguistics (ACL 2022)*

[122] Satyapriya Krishna, Rahul Gupta, Apurv Verma, Jwala Dhamala, Yada Pruksachatkun, and **Kai-Wei Chang**, Measuring Fairness of Text Classifiers via Prediction Sensitivity. *the Annual Meeting of the Association for Computational Linguistics (ACL 2022)*

[123] Fan Yin, Zhouxing Shi, Cho-Jui Hsieh, and **Kai-Wei Chang**, On the Sensitivity and Stability of Model Interpretations. *the Annual Meeting of the Association for Computational Linguistics (ACL 2022)*

[124] Yang Trista Cao, Yada Pruksachatkun, **Kai-Wei Chang**, Rahul Gupta, Varun Kumar, Jwala Dhamala, and Aram Galstyan, On the Intrinsic and Extrinsic Fairness Evaluation Metrics for Contextualized Language Representations. *the Annual Meeting of the Association for Computational Linguistics (ACL 2022, short)*

[125] Jianhan Xu, Cenyuan Zhang, Xiaoqing Zheng, Linyang Li, Cho-Jui Hsieh, **Kai-Wei Chang**, and Xuanjing Huang. Towards Adversarially Robust Text Classifiers by Learning to Reweight Clean Examples. *the Annual Meeting of the Association for Computational Linguistics (ACL-Finding 2022)*

[126] Umang Gupta, Jwala Dhamala, Varun Kumar, Apurv Verma, Yada Pruksachatkun, Satyapriya Krishna, Rahul Gupta, **Kai-Wei Chang**, Greg Ver Steeg, and Aram Galstyan, Mitigating Gender Bias in Distilled Language Models via Counterfactual Role Reversal. *the Annual Meeting of the Association for Computational Linguistics (ACL-Finding 2021)*

[127] Cenyuan Zhang, Xiang Zhou, Yixin Wan, Xiaoqing Zheng, **Kai-Wei Chang**, and Cho-Jui Hsieh, Improving the Adversarial Robustness of NLP Models by Information Bottleneck. *the Annual Meeting of the Association for Computational Linguistics (ACL-Finding 2021)*

[128] Zhecan Wang, Haoxuan You, Liunian Harold Li, Alireza Zareian, Suji Park, Yiqing Liang, **Kai-Wei Chang**, and Shih-Fu Chang, SGEITL: Scene Graph Enhanced Image-Text Learning for Visual Commonsense Reasoning. *AAAI Conference on Artificial Intelligence (AAAI 2022)*.

[129] D. Yin, L. Harold Li, Z. Hu, N. Peng, and **K.-W. Chang**, Broaden the Vision: Geo-Diverse Visual Commonsense Reasoning. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*.

[130] S. Dev, M. Monajatipoor, A. Ovalle, A. Subramonian, J. Phillips, and **K.-W. Chang**, Harms of Gender Exclusivity and Challenges in Non-Binary Representation in Language Technologies. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*.

[131] L. Yuan, X. Zheng, Y. Zhou, C.-J. Hsieh, and **K.-W. Chang**. On the Transferability of Adversarial Attacks against Neural NLP Models *Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*.

[132] Z. Li, J. Xu, J. Zeng, L. Li, X. Zheng, Q. Zhang, **K.-W. Chang**, and C.-J. Hsieh. Searching for an Effiective Defender: Benchmarking Defense against Adversarial Word Substitution. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*.

[133] K.-H. Huang, W. Ahmed, N. Peng, and **K.-W. Chang**, Improving Zero-Shot Cross-Lingual Transfer Learning via Robust Training. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*.

[134] Z. Hu, Y. Sun, and **K.-W. Chang**, Relation-Guided Pre-Training for Open-Domain Question Answering. *Conference on Empirical Methods in Natural Language Processing (EMNLP finding 2021)*.

[135] R. Parvez, W. Ahmad, S. Chakraborty, B. Ray, and **K.-W. Chang**, Retrieval Augmented Code Generation and Summarization. *Conference on Empirical Methods in Natural Language Processing (EMNLP finding 2021)*.

[136] T. Meng and **K.-W. Chang**, An Integer Linear Programming Framework for Mining Constraints from Data. *the 32nd International Conference on Machine Learning (ICML 2021).*

[137] W. Ahmad, X. Bai, S. Lee, and **K.-W. Chang**, Select, Extract and Generate: Neural Keyphrase Generation with Layer-wise Coverage Attention. *the Annual Meeting of the Association for Computational Linguistics (ACL 2021)*

[138] W. Ahmad, J. Chi, T. Le, T. Norton, Y. Tian, **K.-W. Chang**, Intent Classification and Slot Filling for Privacy Policies. *the Annual Meeting of the Association for Computational Linguistics (ACL 2021)*

[139] W. Ahmad, H. Li, **K.-W. Chang**, Yashar Mehdad, Syntax-augmented Multilingual BERT for Cross-lingual Transfer. *the Annual Meeting of the Association for Computational Linguistics (ACL 2021)*

[140] E. Sheng, **K.-W. Chang**, P. Natarajan, N. Peng, Societal Biases in Language Generation: Progress and Challenges *the Annual Meeting of the Association for Computational Linguistics (ACL 2021)*

[141] Y. Zhou, X. Zheng, C.-J. Hsieh, **K.-W. Chang**, X. Huang, Defense against Synonym Substitution-based Adversarial Attacks via Dirichlet Neighborhood Ensemble. *the Annual Meeting of the Association for Computational Linguistics (ACL 2021)*

[142] Y. Pruksachatkun, S. Krishna, J. Dhamala, R. Gupta, **K.-W. Chang**, Does Robustness Improve Fairness? Approaching Fairness with Word Substitution Robustness Methods for Text Classification. *the Annual Meeting of the Association for Computational Linguistics (ACL 2021, Finding)*

[143] J. Zhao, D. Khashabi, T. Khot, A. Sabharwal **K.-W. Chang**. Ethical-Advice Taker: Do Language Models Understand Natural Language Interventions? *the Annual Meeting of the Association for Computational Linguistics (ACL 2021, short, Finding)*

[144] R. Parvez, **K.-W. Chang**. Evaluating the Values of Sources in Transfer Learning. *North American Chapter of the Association for Computational Linguistics (NAACL 2021)*.

[145] E. Sheng, **K.-W. Chang**, P. Natarajan, N. Peng. "Nice Try, Kiddo": Ad Hominems in Dialogue Systems. *North American Chapter of the Association for Computational Linguistics (NAACL 2021)*.

[146] C. Zhang, J. Zhao, H. Zhang, **K.-W. Chang**, C.-J. Hsieh. Double Perturbation: On the Robustness of Robustness and Counterfactual Bias Evaluation. *North American Chapter of the Association for Computational Linguistics (NAACL 2021)*.

[147] L. Li, H. You, Z. Wang, A. Zareian, S.-F. Chang, **K.-W. Chang**. Unsupervised Vision-and-Language Pre-training Without Parallel Images and Captions *North American Chapter of the Association for Computational Linguistics (NAACL 2021)*.

[148] J. Huang, K.-H. Huang, and **K.-W. Chang**. Disentangling Semantics and Syntax in Sentence Embeddings with Pre-trained Language Models. *North American Chapter of the Association for Computational Linguistics (NAACL 2021)*.

[149] A. Uppunda, S. Cochran, J. Foster, A. Arseniev-Koehler, V. Mays, **K.-W. Chang**. Adapting Coreference Resolution for Processing Violent Death Narratives. *North American Chapter of the Association for Computational Linguistics (NAACL 2021)*.

[150] J. Dhamala, T. Sun, V. Kumar, S. Krishna, Y. Pruksachatkun, **K.-W. Chang**, R. Gupta. BOLD: Dataset and metrics for measuring biases in open-ended language generation. *in FAccT*, 2021.

[151] K.-H. Huang and **K.-W. Chang**. Generating Syntactically Controlled Paraphrases without Using Annotated Parallel Pairs. *European Chapter of the Association for Computational Linguistics (EACL 2021)*

[152] Y. Zhou, Y. Yan, R. Han, H. Caufield, **K.-W. Chang**, Y. Sun, P. Ping, W. Wang. Clinical Temporal Relation Extraction with Probabilistic Soft Logic Regularization and Global Inference. *AAAI Conference on Artificial Intelligence (AAAI 2021)*.

[153] W. Ahmad, N. Peng, **K.-W. Chang**. GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction. *AAAI Conference on Artificial Intelligence (AAAI 2021)*.

[154] K. Xu, Z. Shi, H. Zhang, Y. Wang, **K.-W. Chang**, M. Huang, B. Kailkhura, X. Lin, and C.-J. Hsieh, Provable, Scalable and Automatic Perturbation Analysis on General Computational Graphs. *Neural Information Processing Systems (NeurIPS 2020).*

[155] K.-H. Huang, C. Li, **K.-W. Chang**. Generating Sports News from Live Commentary: A Chinese Dataset for Sports Game Summarization. *Asia Chapter of the Association for Computational Linguistics (AACL 2020)*

[156] Z. Hu, Y. Dong, K. Wang, **K.-W. Chang**, and Y. Sun. GPT-GNN: Generative Pre-Training of Graph Neural Networks. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2020).*

[157] E. Sheng, **K.-W. Chang**, P. Natarajan, N. Peng. Towards Controllable Biases in Language Generation. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2020, finding).*

[158] L. Liu, Y. Zhou, J. Xu, X. Zheng, **K.-W. Chang**, X. Huang. Cross-Lingual Dependency Parsing by POS-Guided Word Reordering. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2020, finding).*

[159] J. Zhao and **K.-W. Chang**. LOGAN: Local Group Bias Detection by Clustering. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2020, short).*

[160] W. Ahmad, J. Chi, Y. Tian and **K.-W. Chang**. PolicyQA: A Reading Comprehension Dataset for Privacy Policies. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2020, short finding).*

[161] F. Yin, Q. Long, T. Meng, and **K.-W. Chang**, On the Robustness of Language Encoders against Grammatical Errors. *the Annual Meeting of the Association for Computational Linguistics (ACL 2020)*

[162] D. Yin, T. Meng, and **K.-W. Chang**, SentiBERT: An Effective, Transferable and Interpretable Architecture for Compositional Sentiment Semantics. *the Annual Meeting of the Association for Computational Linguistics (ACL 2020)*

[163] J. Zhao, S. Mukherjee, S. Hosseini, **K.-W. Chang**, and A. H. Awadallah, Gender Bias in Multilingual Embeddings and Cross-Lingual Transfer. *the Annual Meeting of the Association for Computational Linguistics (ACL 2020)*

[164] A. Gaut, T. Sun, S. Tang, Y. Huang, J. Qian, M. ElSherief, J. Zhao, D. Mirza, E. Belding, **K.-W. Chang**, and W. Wang, Towards Understanding Gender Bias in Relation Extraction. *the Annual Meeting of the Association for Computational Linguistics (ACL 2020)*

[165] Y. Zhou, J.-Y. Jiang, J. Zhao, **K.-W. Chang**, and W. Wang, "The Boating Store Had Its Best Sail Ever": Pronunciation-attentive Contextualized Pun Recognition. *the Annual Meeting of the Association for Computational Linguistics (ACL 2020)*

[166] Y. S. Jia, T. Meng, J. Zhao, and **K.-W. Chang**, Mitigating Gender Bias Amplification in Distribution by Posterior Regularization *the Annual Meeting of the Association for Computational Linguistics (ACL 2020, short)*

[167] W. Ahmad, S. Chakraborty, B. Ray, and **K.-W. Chang**, A Transformer-based Approach for Source Code Summarization. *the Annual Meeting of the Association for Computational Linguistics (ACL 2020, short)*

[168] L. Li, M. Yatskar, D. Yin, C.-J. Hsieh, and **K.-W. Chang**, What Does BERT with Vision Look At? *the Annual Meeting of the Association for Computational Linguistics (ACL 2020, short)*

[169] Z. Shi, H. Zhang, **K.-W. Chang**, M. Huang, and C.-J. Hsieh, Robustness Verification for Transformers. *International Conference on Learning Representation (ICLR 2020).*

[170] W. Ahmad, Z. Zhang, X. Ma, **K.-W. Chang**, N. Peng. Cross-lingual Dependency Parsing with Unlabeled Auxiliary Languages. *The SIGNLL Conference on Computational Natural Language Learning (CoNLL 2019).*

[171] M. Chen, Y. Tian, H. Chen, **K.-W. Chang**, S. Skiena, C. Zaniolo. Learning to Represent Bilingual Dictionaries. *The SIGNLL Conference on Computational Natural Language Learning (CoNLL 2019).*

[172] T. Meng, N. Peng, **K.-W. Chang**. Target Language-Aware Constrained Inference for Cross-lingual Dependency Parsing. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2019)*.

[173] Y. Zhou, J.-Y. Jiang, **K.-W. Chang**, W. Wang. Learning to Discriminate Perturbations for Blocking Adversarial Attacks in Text Classification. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2019)*.

[174] P. Zhou, W. Shi, J. Zhao, K.-H. Huang, M. Chen, R. Cotterell, **K.-W. Chang**. Examining Gender Bias in Languages with Grammatical Gender. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2019)*.

[175] R. Parvez, T. Bolukbasi, **K.-W. Chang**, V. Saligrama. Robust Text Classifier on Test-Time Budgets. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2019, short)*.

[176] W. Shi, M. Chen, P. Zhou, **K.-W. Chang**. Retrofitting Contextualized Word Embeddings with Paraphrases. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2019, short)*.

[177] E. Sheng, **K.-W. Chang**, P. Natarajan, N. Peng. The Woman Worked as a Babysitter: On Biases in Language Generation. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2019, short)*.

[178] C. Xia, H. Zhang, J. Moghtader, A. Wu, **K.-W. Chang**. Visualizing Trend of Key Roles in News Articles. *Conference on Empirical Methods in Natural Language Processing. (EMNLP 2019, demo)*.

[179] T. Wang, J. Zhao, M. Yatskar, **K.-W. Chang**, V. Ordonez. Balanced Datasets Are Not Enough: Estimating and Mitigating Gender Bias in Deep Image Representations. *International Conference on Computer Vision (ICCV 2019)*.

[180] W. Ahmad, **K.-W. Chang**, H. Wang. Context Attentive Document Ranking and Query Suggestion. *International ACM SIGIR Conference on Research and Development in Information Retrieval. (SIGIR 2019)*.

[181] Z. Hu, T. Chen, **K.-W. Chang**, Y. Sun. Few-Shot Representation Learning for Out-Of-Vocabulary Words. *the Annual Meeting of the Association for Computational Linguistics (ACL 2019)*.

[182] T. Sun, A. Gaut, S. Tang, Y. Huang, M. ElSherief, J. Zhao, D. Mirza. **K.-W. Chang**, W. Wang. Debiasing Gender in Natural Language Processing: Literature Review. *the Annual Meeting of the Association for Computational Linguistics (ACL 2019)*.

[183] W. Ahmad, Z. Zhang, X. Ma, E. Hovy, **K.-W. Chang**, N. Peng. On Difficulties of Cross-Lingual Transfer with Order Differences: A Case Study on Dependency Parsing. *North American Chapter of the Association for Computational Linguistics (NAACL 2019)*.

[184] J. Zhao, T. Wang, M. Yatskar, R. Cotterell, V. Ordonez, **K.-W. Chang**. Gender Bias in Contextualized Word Embeddings. *North American Chapter of the Association for Computational Linguistics (NAACL 2019, short)*.

[185] M. Chen, C. Ju, G. Zhou, X. Chen, T. Zhang, **K.-W. Chang**, C. Zaniolo, W. Wang. Multifaceted Protein-Protein Interaction Prediction Based on Siamese Residual RCNN. *ISMB 2019*.

[186] J. Zhao, Y. Zhou, Z. Li, W. Wang, **K.-W. Chang** Learning Gender-Neutral Word Embeddings *Conference on Empirical Methods in Natural Language Processing (EMNLP 2018, short)*.

[187] R. Parvez, S. Chakraborty, B. Ray, **K.-W. Chang**. Building Language Models for Text with Named Entities. *the Annual Meeting of the Association for Computational Linguistics (ACL 2018)*.

[188] W. Ahmad, **K.-W. Chang**, Hongning Wang. Intent-aware Query Obfuscation for Privacy Protection in Personalized Web Search *International ACM SIGIR Conference on Research and Development in Information Retrieval. (SIGIR 2018).*

[189] C. Jiang, H.-F Yu, C.-J. Hsieh, **K.-W. Chang**, Learning Word Embeddings for Low-resource Languages by PU Learning. *North American Chapter of the Association for Computational Linguistics (NAACL 2018).*

[190] W. Ahmad, **K.-W. Chang**, A Corpus to Learn Refer-to-as Relations for Nominals. *Language Resources and Evaluation Conference (LREC 2018).*

[191] M. Chen, Y. Tian, **Kai-Wei Chang**, S. Skiena, C. Zaniolo Co-training Embeddings of Knowledge Graphs and Entity Descriptions for Cross-lingual Entity Alignment *the 22nd International Joint Conferences on Artificial Intelligence (IJCAI 2018)*

[192] S. Preum, R. Parvez, **K.-W. Chang**, J. Stankovic, A Corpus of Drug Usage Guidelines Annotated with Type of Advice. *Language Resources and Evaluation Conference (LREC 2018).*

[193] L. Feng, M. Ghasemi, **K.-W. Chang**, U. Topcu. Counterexamples for Robotic Planning Explained in Structured Natural Language International Conference on Robotics and Automation (ICRA 2018).

[194] W. Ahmad, **K.-W. Chang**, H. Wang. Multi-Task Learning for Document Ranking and Query Suggestion. *International Confernece on Learning Representation (ICLR 2018).*

[195] K. Arnold, **K.-W Chang**, A. Kalai, Counterfactual Language Model Adaptation for Suggesting Phrases. *International Joint Conference on Natural Language Processing* (IJCNLP 2017).

[196] T. Bolukbasi, **K.-W. Chang**, Joseph Wang, Venkatesh Saligrama. Structured Prediction with Test-time Budget Constraints. *Thirty-First AAAI Conference on Artificial Intelligence (AAAI 2017).*

[197] K.-W. Chang, H. He, H. Daume III, J. Langford, S. Ross A Credit Assignment Compiler for Joint Prediction. *Neural Information Processing Systems (NeurIPS 2016).*

[198] S. Upadhyay, M. Chang, **K.-W. Chang**, W.-t. Yih, Learning from Explicit and Implicit Supervision Jointly For Algebra Word Problems, *Conference on Empirical Methods in Natural Language Processing (EMNLP 2016).*

[199] **K.-W. Chang**, A. Krishnamurthy, A. Agarwal, H. Daumé III, J. Langford. Learning to search better than your teacher, *the 32nd International Conference on Machine Learning (ICML 2015).*

[200] H. Peng, **K.-W. Chang**, D. Roth. A joint framework for coreference resolution and mention head detection, *The SIGNLL Conference on Computational Natural Language Learning (CoNLL 2015).*

[201] **K.-W. Chang**, S. Upadhyay, G. Kundu and D. Roth Structural learning with amortized inference *The Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015).*

[202] **K.-W. Chang**, W.-t. Yih, B. Yang and C. Meek. Typed tensor decomposition of knowledge bases for relation extraction. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014).*

[203] R. Samdani, **K.-W. Chang**, D. Roth. A discriminative latent variable model for online clustering. *the 31st International Conference on Machine Learning (ICML 2014).* .

[204] **K.-W. Chang**, R. Samdani, D. Roth. A constrained latent variable model for coreference resolution. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2013).*

[205] **K.-W. Chang**, W.-t. Yih, C. Meek. Multi-relational latent semantic analysis. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2013).*

[206] **K.-W. Chang**, V. Srikumar, D. Roth. Multi-core structural SVM training. *European Conference on Machine Learning (ECML 2013).*

[207] **K.-W. Chang**, S. Sundararajan, S. S. Keerthi. Tractable semi-supervised learning of complex structured prediction models. *European Conference on Machine Learning (ECML 2013).*

[208] **K.-W. Chang**, B. Deka, W.-M. H. Hwu, D. Roth. Efficient pattern-based time series classification on GPU. *2012 IEEE 12th International Conference on Data Mining (ICDM 2012).*

[209] **K.-W. Chang** and D. Roth, Selective block minimization for faster convergence of limited memory large-scale linear models. *the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2011).*

[210] H.-F. Yu, C.-J. Hsieh, **K.-W. Chang**, and C.-J. Lin, Large linear classification when data cannot fit in memory, *the 22nd International Joint Conferences on Artificial Intelligence (IJCAI 2011, the Best Paper Track).*

[211] F.-L. Huang, C.-J. Hsieh, **K.-W. Chang**, and C.-J. Lin, Iterative scaling and coordinate descent methods for maximum entropy models, *the 47th Annual Meeting of the Association for Computational Linguistics (ACL 2009, short paper).*

[212] S. S. Keerthi, S. Sundararajan, **K.-W. Chang**, C.-J. Hsieh, and C.-J. Lin, A sequential dual method for large scale multi-class linear SVMs, *the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2008).*

Refereed Journal Publications

[213] Cheng-Yi Li, Kao-Jung Chang, Cheng-Fu Yang, Hsin-Yu Wu, Wenting Chen, Hritik Bansal, Ling Chen, Yi-Ping Yang, Yu-Chun Chen, Shih-Pin Chen, Shih-Jen Chen, Jiing-Feng Lirng, **Kai-Wei Chang**, and Shih-Hwa Chiou. âĂIJTowards a Holistic Framework for Multimodal LLM in 3D Brain CT Radiology Report Generation.âĂİ In Nature Communications, 2025.

[214] A. Arseniev-Koehler, S. D. Cochran, V. M. Mays, **K.-W. Chang**, and J. G. Foster. Integrating topic modeling and word embedding to characterize violent deaths, *Proceedings of the National Academy of Sciences* 2022.

[215] A. Arseniev-Koehler, J. G. Foster, V. M. Mays, **K.-W. Chang**, and S. D. Cochran. Aggression, Escalation, and Other Latent Themes in Legal Intervention Deaths of Non-Hispanic Black and White Men: Results From the 2003 - 2017 National Violent Death Reporting System, *American Journal of Public Health, 2021.*

[216] C.-P. Li, **K.-W. Chang**, Distributed Block-diagonal Approximation Methods for Regularized Empirical Risk Minimization, *Machine Learning Journal, 2019.*

[217] L. Li, P. Chen, C.-J. Hsieh, **K.-W. Chang**, Efficient Contextual Representation Learning With Continuous Outputs, *Transactions of the Association for Computational Linguistics, 2019*

[218] D. Duong, W. Ahmad, E. Eskin, **K.-W. Chang**, J. Li. Word and sentence embedding tools to measure semantic similarity of Gene Ontology terms by their definitions *Journal of Computational Biology 2018*

[219] H.-F. Yu, C.-J. Hsieh, **K.-W. Chang**, and C.-J. Lin, Large linear classification when data cannot fit in memory, *ACM Transactions on Knowledge Discovery from Data (TKDD) 5(4):23, 2012*

[220] G.-X. Yuan, **K.-W. Chang**, C.-J. Hsieh, C.-J. Lin, A comparison of optimization methods for large-scale L1-regularized linear classification. *Journal of Machine Learning Research 11 (JMLR), 3183-3234, 2010.*

[221] Y.-W. Chang, C.-J. Hsieh, **K.-W. Chang**, Michael Ringgaard, and C.-J. Lin, Training and Testing Low-degree Polynomial Data Mappings via Linear SVM *Journal of Machine Learning Research 11 (JMLR), 1471-1490, 2010.*

[222] F.-L. Huang, C.-J. Hsieh, **K.-W. Chang**, and C.-J. Lin, Iterative scaling and coordinate descent methods for maximum entropy models. *Journal of Machine Learning Research 11 (JMLR), 815-848, 2010.*

[223] R.-E. Fan, **K.-W. Chang**, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research 9 (JMLR), 1871-1874, 2008.*

[224] **K.-W. Chang**, C.-J. Hsieh, and C.-J. Lin, Coordinate descent method for large-scale L2-loss linear SVM. *Journal of Machine Learning Research 9 (JMLR), 1369-1398, 2008.*

Refereed Shared Task System Papers

[225] Kareem Ahmed, Tao Li, Thy Ton, Quan Guo, **Kai-Wei Chang**, Parisa Kordjamshidi, Vivek Srikumar, Guy Van den Broeck, and Sameer Singh.PYLON: A PyTorch Framework for Learning with Constraints. *AAAI Conference on Artificial Intelligence (AAAI demo 2022)*.

[226] A. Rozovskaya, **K.-W. Chang**, D. Roth. The Illinois-Columbia System in the CoNLL-2014 Shared Task *Proceedings of the Eighteenth Conference on Computational Natural Language Learning (CoNLL 2014)* **(1st place)**.

[227] A. Rozovskaya, **K.-W. Chang**, M. Sammons, D. Roth. The University of Illinois System in the CoNLL-2013 Shared Task. *Proceedings of the Seventeenth Conference on Computational Natural Language Learning (CoNLL) 2013* **(1st place)**.

[228] X. Cheng, B. Chen, R. Samdani, **K-W. Chang**, Z. Fei, M. Sammons, J. Wieting, S. Roy, C. Wang, and D. Roth, Illinois Cognitive Computation Group UI-CCG TAC 2013 Entity Linking and Slot Filler Validation Systems, *Text Analysis Conference (TAC 2013).*

[229] **K.-W. Chang**, R. Samdani, A. Rozovskaya, M. Sammons and D. Roth, Illinois-Coref: The UI System in the CoNLL-2012 Shared Task *Proceedings of the Sixteenth Conference on Computational Natural Language Learning (CoNLL) 2012* **(4th place in the English closed track).**

[230] **K.-W. Chang**, R. Samdani, A. Rozovskaya, N. Rizzolo, M. Sammons and D. Roth, Inference Protocols for Coreference Resolution. *Proceedings of the Fifteenth Conference on Computational Natural Language Learning (CoNLL) 2011* **(3rd place in the closed track).**

[231] H.-Y. Lo, **K.-W. Chang**, S.-T. Chen, T.-H. Chiang, C.-S. Ferng, C.-J. Hsieh, Y.-K. Ko, T.-T. Kuo, H.-C. Lai, K.-Y. Lin, C.-H. Wang, H.-F. Yu, C.-J. Lin, H.-T. Lin and S.-D. Lin. An ensemble of three classifiers for KDD Cup 2009: expanded linear model, heterogeneous boosting, and selective naive Bayes. *Proceedings of KDD-Cup 2009 competition, vol. 7 of JMLR Workshop and Conference Proceedings, 57-64, 2009.* **(3rd Place out of 400+ submissions in the Slow Track).**

Thesis

[232] K.-W. Chang, Selective algorithms for large-scale classification and structured learning (Ph.D.)

[233] K.-W. Chang, A dual coordinate descent method for large-scale linear SVM (Master)

Workshop Publications and Preprints

[234] Hritik Bansal, Nishad Singhi, Yu Yang, Fan Yin, Aditya Grover, and **Kai-Wei Chang**. "CleanCLIP: Mitigating Data Poisoning Attacks in Multimodal Contrastive Learning." In *ICLR workshop on Reliable and Trustworthy Large Scale Machine Learning, 2023* Best Paper .

[235] Z. Deng, W. Shi, P. Zhou, M. Chen, **K.-W. Chang**. Computational Analysis of French-origin Reborrowing Process for English Loanwords. *ICDM Workshop on Multilingual Cognitive Services*, 2019.

[236] Z. Hu, C. Fan, T. Chen, **K.-W. Chang**, Y. Sun. Pre-Training Graph Neural Networks for Generic Structural Feature Extraction. *ICLR Workshop: Representation Learning on Graphs and Manifolds*, 2019.

[237] W. Shi, M. Chen, Y. Tian, **K.-W. Chang**. Learning Bilingual Word Embeddings Using Lexical Definitions *ACL Representation learning for NLP Workshop 2019*.

[238] S. Upadhyay, **K.-W. Chang**, M. Taddy, A. Kalai, J. Zou. Beyond Bilingual: Multi-senseWord Embeddings using Multilingual Context. *ACL Representation learning for NLP Workshop 2017,* **Best Paper Award**.

[239] K. Arnold, **K.-W Chang**, A Kalai. Learning to Suggest Phrases. *AAAI Workshop on Human-Aware AI Workshop*, 2017.

[240] C.-p. Lee, K.-W. Chang, S. Upadhyay, D. Roth. Distributed Training of Structured SVM. *NeurIPS Workshop on Optimization for Machine Learning*, 2015.

[241] K. -W. Chang, H. Daumé III, J. Langford, S. Ross. Efficient Programmable Learning to Search. *ICML Workshop on Machine Learning System*, 2015.

[242] R. Samdani, **K.-W. Chang**, D. Roth. A Discriminative Latent Variable Model for Clustering of Streaming Data with Application to Coreference Resolution. *ICML workshop on Inferning: Interactions between Inference and Learning*, 2013.

[243] H.-F. Yu, C.-J. Hsieh, **K.-W. Chang**, and C-J. Lin, Pascal Challenge: Linear Support Vector Machines. *Pascal Large Scale Learning Challenge in ICML 2008 Workshop*, 2008.

Patents

[244] Efficient polynomial mapping of data for use with linear support vector machines, Y.-W. Chang, C.-J. Hsieh, **K.-W. Chang**, M. Ringgaard, C.-J. Lin, 2013.

[245] Interactive Context-Based Text Completions, K. Arnold, **K.-W Chang**, A Kalai, 2016 (under review).

TALKS

Keynote Talks

- AAAI Workshop on Artificial Intelligence with Biased or Scarce Data, 2024
- NAACL Workshop on Gender Bias in NLP, 2022
- Open Data Science Conference (ODSC) West, 2021.
- West Coast NLP Summit, 2018.
- Southern California Natural Language Processing Symposium, 2017.
- NeurIPS workshop on learning high dimensions with structure, Dec 2016.
- Mid-Atlantic Student Colloquium on Speech, Language and Learning, Breakout Session, Johns Hopkins University, January 2015.

Selected Invited Talks

- Multimodal Representation Learning for Vision and Language
  Baidu 22, Two-Sigma 22
- Unified Pre-Training for Program Understanding and Generation
  Amazon 21, NAACL workshop
- Bias in Language Generation
  Amazon 21.
- Robustness in NLP
  Amazon 20.

- What It Takes to Control Societal Bias in Natural Language Processing
  USC-ISI 20, UCSD 20, UIUC 21, Stanford 18, CMU 18, MSR-NE 19.
- Inject Expert Knowledge and Corpus-Level Constraints in Natural Language Processing Models
  UCLA Stat 19
- Structured Predictions: Practical Advancements and Applications in Natural Language Processing
  UCSD 17, USC 17, Utah 17, Appier Inc. 17, NTU 17, UCLA 17, UMass 17, UCDavis 17
- Multi-Relational Latent Semantic Analysis by Tensor Decomposition
  UMass Med School 16
- Practical Learning Algorithms for Structured Prediction Models
  UMass 15, TTIC 15, UVirginia 15, OSU 15, CMU 15, WSU 15, MSR 15, UArizona 15, MSR-NE 15, Purdue 14, UMD 14, Columbia 14, UIUC 14.