

Dongyeop Kang

412-736-4873 ◊ dongyeop@umn.edu ◊ [dykang.github.io](https://github.com/dykang)

259 Shepherd, 200 Union Street SE, Minneapolis, MN, 55455

Last updated on August 3, 2024

PROFESSIONAL EMPLOYMENT HISTORY

- 2021 – present: Assistant Professor, University of Minnesota, Twin Cities
- 2020 – 2021: Postdoctoral Scholar – University of California, Berkeley (PI: Marti A. Hearst)
- 2015 – 2020: Research Assistant – Carnegie Mellon University (PI: Eduard Hovy)
- 2018 Summer: Research Intern – Facebook AI Research, Menlo Park, CA
- 2017 Summer: Research Intern – Allen Institute for Artificial Intelligence (AI2), Seattle, WA
- 2016 Summer: Research Intern – Microsoft Research, Redmond, WA
- 2014 – 2015: Researcher (alternative military service) – Naver Labs, Seongnam, Korea
- 2012 – 2014: Researcher (alternative military service) – KAIST Institute, Daejeon, Korea
- 2014 Summer: Research Intern – Oracle Labs, Redwood City, CA
- 2013 Summer: Research Intern – IBM T. J. Watson, Yorktown, NY
- 2011: Visiting Researcher – Hong Kong University of Science and Technology, Hong Kong
- 2010 Summer: Research Intern – Microsoft Research, Redmond, WA
- 2009 – 2010: Research Intern – Microsoft Research Asia, Beijing, China

EDUCATION

Carnegie Mellon University

2020

Ph.D., Language Technology Institute, School of Computer Science

Thesis: *Linguistically Informed Language Generation: A Multifaceted Approach*

Advisor: Eduard Hovy

Committee: Jeffrey Bigham, Alan W Black, Jason Weston, Dan Jurafsky

Korea Advanced Institute of Science and Technology (KAIST)

2010

B.S. and M.S., Computer Science Engineering

AWARDS AND HONORS

- 2024 Open Philanthropy Gift award
- 2022-2024 Grammarly Research Gift Award
- 2023 Cisco Research Award
- 2022-2024 3M Non-tenured Faculty Award (NTFA)
- 2022 Best paper award, In2Writing workshop at ACL
- 2022 Sony Research Faculty Innovation Award
- 2018 Top Reviewers, NeurIPS and ACL
- 2018 Allen Institute for AI Research Fellowship
- 2016 CMU Presidential Fellowship
- 2013 SMBA Startup Investment Prize
- 2011 ILJU Foundation Ph.D. Scholarship
- 2007 Summa Cum Laude, KAIST

GRANTS AND RESEARCH GIFTS

- 2024 Gift from Grammarly Research, PI (\$150,000)
- 2024 Open Philanthropy, Lawyer Benchmarks, PI (w Daniel Schwarcz, Brett McDonnell) (\$74,132)

- 2024 UMN Data Science Institute (DSI), Small Seed Grant, PI (w Mingyi Hong), (\$15,000) “Aligning AI models with Human Values: Formulation, Algorithms and Applications”
- 2022 - 2024 Gift from 3M, PI (\$4,500), “Human-centric Natural Language Processing”
- 2023 Computing Grant from OpenAI (\$25,000)
- 2023 UMN Clinical and Translational Science Institute (CTSI), Co-PI (/w Jennifer Needle, Gwenth Fischer) (\$50,000), “AI Assistant for Non-English Speaking Clinical Trial Participants”
- 2023 Computing Grant from Cohere AI (\$5,000), Google Cloud (\$6,000), Oracle Cloud (\$1,000), and Amazon Web Service (\$5,000)
- 2023 Grant from Cisco Research, PI (\$175,000), “Active Instruction Tuning for Efficient and Collaboration Task Annotation”
- 2023 Gift from Accenture Labs, PI (\$85,000), “Writing Assistant for Professionals using NLP+HCI Approaches”
- 2023 Gift from UMN Early Innovation Fund, PI (\$10,431), “Collecting Human Explanations: Cognitively-inspired NLP Systems”
- 2023 MnRI Seed Grant, PI: Karthik Desingh, Co-PI: Dongyeop Kang (\$40,000)
- 2022 Gift from Grammarly Research, PI (\$140,000), “Improving Writing Assistants with Planning and Iterative Revision”
- 2022 Grant from Sony Research, PI (\$100,000), “Goal-oriented Decoding Control in Language Generation”
- 2021 - 2024 Setup grant from University of Minnesota

SERVICE AND PUBLIC OUTREACH

Program Chairs and Organizers

- Founder and Co-organizer, Workshop on “[Pluralistic Alignment](#)” at NeurIPS 2024
- Steering committee, Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual ([CustomNLP4U](#)), at EMNLP 2024
- Senior PC, Association for the Advancement of Artificial Intelligence (AAAI) 2023, 2024
- Area Chair (Machine Learning in NLP), Conference on Empirical Methods in Natural Language Processing (EMNLP) 2023
- Founder and Co-organizer, Workshop on “Intelligent and Interactive Writing Assistants ([In2Writing](#))” at ACL 2022, CHI 2023, and CHI 2024
- Co-organizer, [Minnesota NLP Seminar](#), 2021 Fall - 2022 Fall
- Founder and Co-organizer, First Workshop on “[CtrlGen](#): Controllable Generative Modeling in Language and Vision” at NeurIPS 2021

Program Committees (Conferences)

- 2021-2022, ACL Rolling Review
- 2018-Present, ACL (Association for Computational Linguistics)
- 2019, 2022, NAACL (Annual Conference of the North American Chapter of the Association for Computational Linguistics)
- 2018-Present, EMNLP (Conference on Empirical Methods in Natural Language Processing)
- 2018-Present, NeurIPS (Conference on Neural Information Processing Systems)
- 2019-Present, ICLR (International Conference on Learning Representations)
- 2019, ICML (International Conference on Machine Learning)

Program Committees (Workshops and Journals)

- 2018, MRQA (Machine Reading for Question Answering) workshop
- 2019, WNUT (Workshop on Noisy User-generated Text) workshop
- 2019, Scientometrics (International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy)

Department and University Service

- 2021-2023, Admissions Committee, University of Minnesota
- 2023-Present, Curriculum Committee, University of Minnesota

PUBLICATIONS

Dissertations

- D1 Dongyeop Kang, “Linguistically Informed Language Generation: A Multifaceted Approach”, Committees: Edward Hovy, Jeffrey Bigham, Alan W Black, Jason Weston, Dan Jurafsky, PhD Dissertation 2020

Archived Preprints and Papers Under Review

- A11 Daechul Ahn, Yura Choi, San Kim, Youngjae Yu, Dongyeop Kang, Jonghyun Choi “i-SRT: Aligning Large Multimodal Models for Videos by Iterative Self-Retrospective Judgment” <https://arxiv.org/abs/2406.11280v1> under review
- A10 Shirley Anugrah Hayati, Taehee Jung, Tristan Bodding-Long, Sudipta Kar, Abhinav Sethy, Joo-Kyung Kim, Dongyeop Kang, “Chain of Instructions: Compositional Instruction Tuning on Large Language Models”, <https://arxiv.org/abs/2402.11532> under review
- A9 Anna Martin-Boyle, Aahan Tyagi, Marti A. Hearst, Dongyeop Kang, “Shallow Synthesis of Knowledge in GPT-Generated Texts: A Case Study in Automatic Related Work Composition”, <https://arxiv.org/abs/2402.12255> under review
- A8 Ritik Sachin Parkar, Jaehyung Kim, Jong Inn Park, Dongyeop Kang, “SelectLLM: Can LLMs Select Important Instructions to Annotate?” <https://arxiv.org/abs/2401.16553> under review
- A7 Karin de Langis, Ryan Koo, Dongyeop Kang, “Reinforcement Learning with Dynamic Multi-Reward Weighting for Multi-Style Controllable Generation”, under review
- A6 Debarati Das*, Karin de Langis*, Anna Martin-Boyle*, Jaehyung Kim*, Minhwa Lee*, Zae Myung Kim*, Shirley Anugrah Hayati, Risako Owan, Bin Hu, Ritik Sachin Parkar, Ryan Koo, Jong Inn Park, Aahan Tyagi, Libby Ferland, Sanjali Roy, Vincent Liu, Dongyeop Kang, “Under the Surface: Tracking the Artifactuality of LLM-Generated Data”, <https://arxiv.org/abs/2401.14698> under review
- A5 Tianhao Zhang, Zhexiong Lin, Zhecheng Sheng, Chen Jiang, Dongyeop Kang “On the Sequence Evaluation based on Stochastic Processes” <https://arxiv.org/abs/2405.17764> under review
- A4 Chenliang Li, Zeyi Liao, Siliang Zeng, Dongyeop Kang, Alfredo Garcia, Mingyi Hong, “Joint Demonstration and Preference Learning Improves Policy Alignment with Human Feedback”, <https://arxiv.org/abs/2406.06874> under review
- A3 Shirley Anugrah Hayati, Minhwa Lee, Dheeraj Rajagopal, Dongyeop Kang, “How Far Can We Extract Diverse Perspectives from Large Language Models?” <https://arxiv.org/abs/2311.09799> under review
- A2 Hao Zou, Zae Myung Kim, Dongyeop Kang, “Diffusion Models in NLP: A Survey” <https://arxiv.org/abs/2305.14671> under review
- A1 Anna Martin-Boyle, Andrew Head, Kyle Lo, Risham Sidhu, Marti A Hearst, Dongyeop Kang, “Complex Mathematical Symbol Definition Structures: A Dataset and Model for Coordination Resolution in Definition Extraction” <https://arxiv.org/abs/2305.14660> under review

Refereed Conference Proceedings

- C52 Haechan Kim, Junho Myung, Seoyoung Kim, Sungpah Lee, Dongyeop Kang, Juho Kim “LearnerVoice: A Dataset of Non-Native English Learners’ Spontaneous Speech” Interspeech 2024
- C51 Carl Winge, Adam Imdieke, Bahaa Aldeeb, Dongyeop Kang, Karthik Desingh, “Talk Through It: End User Directed Manipulation Learning” IEEE Robotics and Automation Letters (RA-L) 2024
- C50 Zae Myung Kim, Kwang Hee Lee, Preston Zhu, Vipul Raheja, Dongyeop Kang, “Threads of Subtlety: Detecting Machine-Generated Texts Through Discourse Motifs” Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2024 [link](#)
- C49 Daechul Ahn, Yura Choi, Youngjae Yu, Dongyeop Kang, Jonghyun Choi, “Tuning Large Multimodal Models for Videos using Reinforcement Learning from AI Feedback ”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2024 [link](#) [Oral](#)
- C48 Jihyung Kil, Joo-Kyung Kim, Dongyeop Kang, “II-MMR: Identifying and Improving Multi-modal Multi-hop Reasoning in Visual Question Answering”, Findings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2024 [link](#)
- C47 Ryan Koo, Minhwa Lee, Vipul Raheja, Jong Inn Park, Zae Myung Kim, Dongyeop Kang, “Benchmarking Cognitive Biases in Large Language Models as Evaluators”, Findings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2024 [link](#)

- C46 Gabriel Garlough-Shah, Jong Inn Park, Shirley Anugrah Hayati, Dongyeop Kang, Jisu Huh, “Consumer Engagement With AI-Powered Search Engines and Implications for the Future of Search Advertising” Association for Education in Journalism and Mass Communication (AEJMC) 2024
- C45 Debarati Das, Ishaan Gupta, Jaideep Srivastava, Dongyeop Kang, “Which Modality should I use – Text, Motif, or Image? : Understanding Graphs with Large Language Models” Findings of the North American Chapter of the Association for Computational Linguistics (NAACL) 2024
- C44 Smitha Muthya Sudheendra, Maral Abdollahi, Dongyeop Kang, Jisu Huh, Jaideep Srivastava, “SkOTaPA: A dataset for Skepticism Detection in Online Text after Persuasion Attempt” The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING) 2024
- C43 Zhecheng Sheng, Tianhao Zhang, Chen Jiang, Dongyeop Kang, “BBScore: A Brownian Bridge Based Metric for Assessing Text Coherence” Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI) 2024
- C42 Ryan Koo, Yekyung Kim, Dongyeop Kang, Jaehyung Kim “Meta-Crafting : Improved Detection of OOD Texts via Crafting Metadata Space” Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI) 2024 (Student Abstract)
- C41 Hyunjoo Chae, Yongho Song, Kai Tzu-iunn Ong, Taeyoon Kwon, Minjin Kim, Youngjae Yu, Dongha Lee, Dongyeop Kang, Jinyoung Yeo, “Dialogue Chain-of-Thought Distillation for Commonsense-aware Conversational Agents” Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023
- C40 Vipul Raheja, Dhruv Kumar, Ryan Koo, Dongyeop Kang, “CoEdit: Text Editing by Task-specific Instruction Tuning” Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023, Media Coverage ([MarTechPost](#)) **10K+ download in HuggingFace since May 2023**
- C39 Karin de Langis, Dongyeop Kang, “A Comparative Study on Textual Saliency of Styles from Eye Tracking, Annotations, and Language Models”, The SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2023
- C38 Risako Owan, Maria Gini, Dongyeop Kang, “Quirk or Palmer: A Comparative Study of Modal Verb Frameworks with Annotated Datasets”, The SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2023
- C37 Daechul Ahn, Daneul Kim, Gwangmo Song, Seung Hwan Kim, Honglak Lee, Dongyeop Kang, Jonghyun Choi “Story Visualization by Online Text Augmentation with Context Memory”, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2023
- C36 Katie Kim, Bugil Chang, Debarati Das, Smitha Muthya, Jisu Huh, Dongyeop Kang, Jaideep Srivastava “Rebuilding Social Connection and Enhancing Advertising Effects Through the Nostalgic Appeal during the Pandemic”, Association for Education in Journalism and Mass Communication (AEJMC) 2023
- C35 Jaehyung Kim, Yekyung Kim, Karin De Langis, Jinwoo Shin, and Dongyeop Kang “A Universal Framework for Dataset Characterization with Multidimensional Meta-information”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2023
- C34 Haneul Yoo, Rifki Afina Putri, Changyoon Lee, Youngin Lee, So-Yeon Ahn, Dongyeop Kang, Alice Oh “Rethinking Annotation: Can Language Learners Contribute?”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2023
- C33 Debarati Das, David Ma, Dongyeop Kang “Balancing Effect of Training Dataset Distribution of Multiple Styles for Multi-Style Text Transfer”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) Findings 2023
- C32 Jaehyung Kim, Jinwoo Shin, Dongyeop Kang “Prefer to Classify: Improving Text Classifiers via Auxiliary Preference Learning”, International Conference on Machine Learning (ICML) 2023
- C31 Taehee Jung, Joo-Kyung Kim, Sungjin Lee, Dongyeop Kang, “Cluster-Guided Label Generation in Extreme Multi-Label Classification”, Conference of the European Chapter of the Association for Computational Linguistics (EACL) 2023
- C30 Shirley Anugrah Hayati, Kyumin Park, Dheeraj Rajagopal, Lyle Ungar, Dongyeop Kang, “StyLEx: Explaining Styles with Lexicon-Based Human Perception”, Conference of the European Chapter of the Association for Computational Linguistics (EACL) 2023
- C29 Ruyuan Wan, Jaehyung Kim, Dongyeop Kang. “Everyone’s Voice Matters: Quantifying Annotation Disagreement Using Demographic Information”, Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI) 2023
Oral

- C28 Kyle Lo, Joseph Chee Chang, Andrew Head, Jonathan Bragg, Amy X. Zhang, Cassidy Trier, Chloe Anastasiades, Tal August, Russell Authur, Danielle Bragg, Erin Bransom, Isabel Cachola, Stefan Candra, Yoganand Chandrasekhar, Yen-Sung Chen, Evie Yu-Yen Cheng, Yvonne Chou, Doug Downey, Rob Evans, Raymond Fok, Fangzhou Hu, Regan Huff, Dongyeop Kang, Tae Soo Kim, Rodney Kinney, Aniket Kittur, Hyeonsu Kang, Egor Klevak, Bailey Kuehl, Michael Langan, Matt Latzke, Jaron Lochner, Kelsey MacMillan, Eric Marsh, Tyler Murray, Aakanksha Naik, Ngoc-Uyen Nguyen, Srishti Palani, Soya Park, Caroline Paulic, Napol Rachatasumrit, Smita Rao, Paul Sayre, Zejiang Shen, Pao Siangliulue, Luca Soldaini, Huy Tran, Madeleine van Zuylen, Lucy Lu Wang, Christopher Wilhelm, Caroline Wu, Jiangjiang Yang, Angele Zamarron, Marti A. Hearst, Daniel S. Weld. “The Semantic Reader Project: Augmenting Scholarly Documents through AI-Powered Interactive Reading Interfaces” Communications of ACM (CACM) 2023
- C27 Zae Myung Kim, Wanyu Du, Vipul Raheja, Dhruv Kumar, Dongyeop Kang. “Improving Iterative Text Revision by Learning Where to Edit from other Revision Tasks”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022
- C26 Wanyu Du, Vipul Raheja, Dhruv Kumar, Zae Myung Kim, Melissa Lopez, Dongyeop Kang. “Understanding Iterative Revision from Human-Written Text.”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2022
- C25 Jaehyung Kim, Dongyeop Kang, Sungsoo Ahn, Jinwoo Shin. “What Makes Better Augmentation Strategies? Augment Difficult but Not too Different”, International Conference on Learning Representations (ICLR) 2022
- C24 Wanyu Du*, Zae Myung Kim*, Vipul Raheja, Dhruv Kumar, Dongyeop Kang (*equal contribution), “Read, Revise, Repeat: A System Demonstration for Human-in-the-loop Iterative Text Revision”, In2Writing @ACL 2022, **Best paper award**
- C23 Hayati, Shirley Anugrah, Dongyeop Kang and Lyle Ungar. “Does BERT Learn as Humans Perceive? Understanding Linguistic Styles through Lexica.”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021
- C22 Hwiyeol Jo, Dongyeop Kang, Andrew Head, Marti A. Hearst, “Modeling Mathematical Notation Semantics in Academic Papers ”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) Findings, 2021
- C21 Jinwoo Nam, Daechul Ahn, Dongyeop Kang, Seong Jong Ha, Jonghyun Choi, “Zero-Shot Natural Language Video Localization”, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2021 **Oral** (Top 3%)
- C20 Dongyeop Kang, Eduard Hovy, “Style is NOT a single variable: Case Studies for Cross-Stylistic Language Understanding”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2021 **Oral**
- C19 Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, Marti A. Hearst, “Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols”, ACM Conference on Human Factors in Computing Systems (CHI) 2021
- C18 Dongyeop Kang, Eduard Hovy, “Plan ahead: Self-Supervised Text Planning for Paragraph Completion Task”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020 **Oral**
- C17 Zae Myung Kim, Vassilina Nikoulina, Dongyeop Kang, Didier Schwab, Laurent Besacier. “Visualizing Cross-Lingual Discourse Relations in Multilingual TED Corpora”, CODI @EMNLP 2021
- C16 Dyah Adila, Dongyeop Kang, “Understanding Out-of-distribution: A Perspective of Data Dynamics”, ICBINB @NeurIPS 2021
- C15 Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, Zhou Yu, “INSPIRED: Toward Sociable Recommendation Dialog Systems”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020
- C14 Taehee Jung, Dongyeop Kang, Hua Cheng, Lucas Mentch, Thomas Schaaf, “Posterior Calibrated Training on Sentence Classification Tasks”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2020
- C13 Dongyeop Kang*, Taehee Jung*, Lucas Mentch, Eduard Hovy (*equal contribution), “Earlier Isn’t Always Better: Sub-aspect Analysis on Corpus and System Biases in Summarization”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019
- C12 Dongyeop Kang, Anusha Balakrishnan, Pararth Shah, Paul Crook, Y-Lan Boureau, Jason Weston, “Recommendation as a Communication Game: Self-Supervised Bot-Play for Goal-oriented Dialogue”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019

- C11 Dongyeop Kang, Varun Gangal, Eduard Hovy, “(Male, Bachelor) and (Female, Ph.D) have different connotations: Parallely Annotated Stylistic Language Dataset with Multiple Personas”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019 **Oral**
- C10 Dongyeop Kang, Hiroaki Hayashi, Alan W Black, Eduard Hovy, “Linguistic Versus Latent Relations for Modeling Coherent Flow in Paragraphs”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019
- C9 Dongyeop Kang, Tushar Khot, Ashish Sabharwal, Peter Clark, “Bridging Knowledge Gaps in Neural Entailment via Symbolic Models”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2018
- C8 Dongyeop Kang, Tushar Khot, Ashish Sabharwal, Eduard Hovy, “AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples”, Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2018
- C7 Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Madeleine van Zuylen, Sebastian Kohlmeier, Eduard Hovy, Roy Schwartz, “A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications”, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL) 2018
- C6 Chu-Cheng Lin, Dongyeop Kang, Michael Gamon, Madian Khabsa, Ahmed Hassan Awadallah, Patrick Pantel, “Actionable email intent modeling with reparametrized RNN”, Association for the Advancement of Artificial Intelligence (AAAI) 2018
- C5 Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, Eduard Hovy, “Detecting and Explaining Causes From Text For a Time Series Event”, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2017
- C4 Dongyeop Kang, DongGyun Han, Na Hea Park, Sangtae Kim, U Kang, Soobin Lee, “Eventera: Real-time Event Recommendation System from Massive Heterogeneous Online Media”, ICDM 2014
- C3 Dongyeop Kang, Woosang Lim, Kijung Shin, Sael Lee, U Kang, “Data/Feature Distributed Stochastic Coordinate Descent for Logistic Regression”, Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM), 2014
- C2 Dongyeop Kang, Youngja Park, Suresh Chari, “Hetero-Labeled LDA: A partially supervised topic model with heterogeneous label information”, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD) 2014
- C1 Dongyeop Kang, Daxin Jiang, Jian Pei, Zhen Liao, Xiaohui Sun, Ho-Jin Choi, “Multidimensional Mining of Large-Scale Search Logs: A Topic-Concept Cube Approach”, ACM International Conference on Web Search and Data Mining (WSDM) 2011

Refereed Workshop, Posters, Extended Abstracts, and Demos

- W9 Minhwa Lee, Zae Myung Kim, Vivek Khetan, Dongyeop Kang, “Human-AI Collaborative Taxonomy Construction: A Case Study in Profession-Specific Writing Assistants” In2Writing @ CHI 2024
- W8 Ruixin Yang, Dheeraj Rajagopal, Shirley Anugrah Hayati, Bin Hu, Dongyeop Kang, “Confidence Calibration and Rationalization for LLMs via Multi-Agent Deliberation” Workshop on Reliable and Responsible Foundation Models, ICLR 2024
- W7 London Lowmanstone*, Ruyuan Wan*, Risako Owan, Jaehyung Kim, Dongyeop Kang “Annotation Imputation to Individualize Predictions: Initial Studies on Distribution Dynamics and Model Predictions”, Workshop on Perspectivist Approaches to NLP (NLPerspectives) @ECAI 2023
- W6 Rose Neis, Karin de Langis, Zae Myung Kim, Dongyeop Kang “An Analysis of Reader Engagement in Literary Fiction through Eye Tracking and Linguistic Features”, Workshop on Narrative Understanding (WNU) @ACL 2023
- W5 Ryan Hyunkyo Koo*, Anna Martin*, Linghe Wang, Dongyeop Kang, “Decoding the End-to-end Writing Trajectory in Scholarly Manuscripts”, In2Writing @CHI 2023
- W4 Ruyuan Wan, Naome Etori, Karla Badillo-Urquiola, Dongyeop Kang, “User or Labor: An Interaction Framework for Human-Machine Relationships in NLP”, Workshops on Data Science with Human in the Loop (DaSH) @EMNLP 2022
- W3 Steven Y. Feng*, Varun Gangal*, Dongyeop Kang, Teruko Mitamura, Eduard Hovy, “GenAug: Data Augmentation for Finetuning Text Generators”, DeeLIO Workshop @EMNLP 2020

- W2 Dongyeop Kang, Andrew Head, Risham Sidhu, Kyle Lo, Daniel Weld, Marti A. Hearst, “Document-Level Definition Detection in Scholarly Documents: Existing Models, Error Analyses, and Future Directions”, SDP Workshop @EMNLP 2020
- W1 Jung-Woo Ha, Dongyeop Kang, Hyuna Pyo, Jeonghee Kim, “News2Images: Automatically Summarizing News Articles into Image-Based Contents via Deep Learning”, INRA Workshop @RecSys 2015

INVITED INVITED SPEAKERS

Invited Speakers and Lectures

- Invited Talk, *Under the Surface: Tracking the Artifactuality of LLM-Generated Data*, Megagon Labs, May 2024
- Guest lecture, *Human-centric NLP in era of Large Language Models*, JOUR8500 - Advanced Methods Seminar: Computational Communication Research, UMN, Mar 2024
- Invited speaker, *Scaffolding AI with Human Cognition*, Thomson Reuters Labs, Nov 2023
- Invited guest lecturer, *AI for Scholarly Reading and Writing*, Pennsylvania State University (PSU), Sept 2023
- Invited Speaker, *Human-centric Language Technologies: Of the Human, By the Human, For the Human*, UMN NLP/IE group, Sept 2023
- Invited Speaker, *Human-centric AI: For the Human, Of the Human, by the Human*, 3M NTFA Symposium, Aug 2023
- Invited Speaker, *Disagreement Modeling in the era of Large Language Models*, Microsoft Research, May 2023
- Invited Panelist, *ChatGPT and the Future of AI-Assisted Professionals: A Discussion with UMN Law, Computer Science, and Education Experts on How AI is Disrupting Professional and Educational Spaces*, UMN Law School, May 2023
- Interview with Echo (student-run news site of St. Louis Park High School) on future of AI in classroom (GPT or GPA), Apr 2023
- Invited Speaker, *Career Paths as an AI researcher*, UMN Korean-American Scientists and engineers association (KSEA), Mar 2023
- Invited Speaker, *Modeling disagreement among humans for inclusive NLP systems*, Google People+AI Research (PAIR), Feb 2023
- Invited Speaker, *Human-centric NLP*, Hyundai AI, Dec 2022
- Invited Speaker, *Human-centric NLP*, NSF ROSE-HUB, Nov 2022
- Invited Speaker, *Human-centric NLP*, 3M NTFA Symposium, Sep 2022
- Invited Speaker, *Human-centric NLP: Learning from Human Perceptions, Interactions, and Disagreements*, Yonsei University, Jul 2022
- Invited Speaker, *Human-centric NLP: Learning from Human Perceptions, Interactions, and Disagreements*, KAIST CS, Jul 2022
- Invited Speaker, *Human-centric NLP: Learning from Human Perceptions, Interactions, and Disagreements*, LG AI Research, Jul 2022
- Invited Speaker, *Human-centric NLP: Learning from Human Perceptions, Interactions, and Disagreements*, Samsung Research, Jul 2022
- Invited Speaker, *Human-centric NLP: Learning from Human Perceptions, Interactions, and Disagreements*, Grammarly, Jul 2022
- Invited Speaker, *Human-centric NLP: Learning from Human Perceptions, Interactions, and Disagreements*, Thomson Reuters Labs, Jul 2022
- Invited Speaker, *Fixing the NLP Pipeline with Humans and Data*, USC ISI AI Seminar, Mar 2022
- Guest lecture, *Introduction to Natural Language Processing*, Intro to Undergraduate Research in CS, UMN, Mar 2022
- Invited Speaker, *Fixing the NLP pipeline with humans and data*, Naver AI US, Feb 2022
- Invited Speaker and panelist, *Language generation: current trends and the road ahead*, CMU LTI Seminar, Feb 2022
- Invited Speaker, *Humans and Data in the loop of NLP Pipeline*, GaTech NLP Seminar, Dec 2021
- Invited Speaker, *Humans and Data in the loop of NLP Pipeline*, Minnesota Robotic Institute Colloquium, Nov 2021
- Invited Speaker, *Humans and Data in the loop of NLP Pipeline*, KAIST SE Colloquium, Nov 2021
- Invited Speaker, *Styles and Interactions in Human-centered NLP*, Naver Labs Europe, Nov 2021
- Invited Speaker, *Styles and Interactions in Human-centered NLP*, Grammarly AI, Aug 2021

- Invited Speaker, *Styles and Interactions in Human-centered NLP*, SNU Summer AI school, Aug 2021
- Invited Speaker, *Toward Human-centric Language Generation Systems*, University of Wisconsin - Madison, CS Department, March 2021
- Invited Speaker, *Toward Human-centric Language Generation Systems*, University of Rochester, CS Department, March 2021
- Invited Speaker, *Toward Human-centric Language Generation Systems*, University of Minnesota - Twin Cities, CSE Colloquium, March 2021
- Invited Speaker, *Toward Human-centric Language Generation Systems*, KAIST, AI Department, Feb 2021
- Invited Speaker, *Linguistically Informed Language Generation*, KAIST AI Colloquium, Nov 2020
- Invited Speaker, *Linguistically Informed Language Generation*, GIST EECS Colloquium, Nov 2020
- Invited Speaker, *Uncertainty Calibration in Text Classifications*, Berkeley AI Research (BAIR) Workshop, July 2020
- Invited Speaker, *Linguistically Informed Language Generation*, CMU LTI Seminar, July 2020
- Invited Speaker, *Linguistically Informed Language Generation*, POSTECH AI Seminar, July 2020

TEACHING

Courses

- CSCI 5541 - Natural Language Processing - Spring 2023, Fall 2023, Spring 2024
- CSCI 5980 - Natural Language Processing (NLP) with Deep Learning - Fall 2022
- CSCI 8980 - Introduction to Natural Language Processing (NLP) Research - Spring 2022

Course Evaluations

Course	When	Num. of Students	Well Prepared	Clear Presentation	Provided Feedback	Treated Me Respectfully	Recommend Instructor?
CSCI 8980	Spring 2022	24	5.7	5.7	5.3	5.8	5.7
CSCI 5541	Spring 2023	59	5.7	5.3	5.5	5.9	5.4
CSCI 5541	Fall 2023	60	5.9	5.6	5.8	5.9	5.7

Table 1: Score range 1-6. 8980 is *Introduction to NLP Research*, 5541 is *Natural Language Processing*

ADVISING AND MENTORING

Current Ph.D. Students

- Karin de Langis, DDF Fellow
- Debarati Das (co-advised with Jaideep Srivastava)
- Risako Owan (co-advised with Maria Gini)
- Zae Myung Kim
- Anna Martin-Boyle, CS&E Fellow
- Shirley Anugrah Hayati
- James Mooney

Current Masters Students

- Ryan Koo
- Linghe Wang
- Wen Jianing

Current Undergraduate Students

- Bin Hu
- Tripp Dow
- Sanjali Roy

- William Walker
- Preston Zhu
- Vincent Liu

Past Ph.D./Masters/Undergraduate Students

- Minhwa Lee, visiting Master student (UMass) → Research Assistant, UMass Lowell
- Ishaan Gupta, CS&E Undergraduate
- Ruixin Yang, visiting Undergraduate (UBC) → GaTech Masters
- Jong Inn Park, Data Science Masters, Plan B → Research Assistant, UMN
- Ritik Parkar, Data Science Masters, Plan B
- Aahan Tyagi, Data Science Masters, Plan B
- Amrutha Shetty, Data Science Masters, Plan B → Data analyst, Malec Engineering
- London Lowmanstone, CS&E Masters → UMN CS Ph.D.
- Hao Zou, CS&E Undergraduate
- Kelsey Neis, Data Science Masters, Plan B
- James Mooney, CS&E Masters → UMN CS Ph.D.
- Ruyuan Wan, Linguistics Masters → University of Notre Dame, CS Ph.D. → Penn State University, CS Ph.D.
- Miguel Miguélez Díaz, Data Science Masters → Amazon
- Yang He, CS&E Masters → Google
- David Ma, CS&E Masters → Optiver
- Jaehyung Kim, Visiting Ph.D. from KAIST → CMU Postdoc → Assistant Professor, Yonsei AI
- Petros Karypis, CS&E Undergraduate → UCSD Masters → UCSD PhD
- Wanyu Du, Grammarly intern → University of Virginia → Amazon AWS

Ph.D./Masters/Undergraduate Committee Member

- Basel Hussein (Ph.D., Learning Technologies, Education & Human Development), *Navigating Learning and Goal Trajectories During Impasses*, current
- Tianhao Zhang (Ph.D., Applied Mathematics), current
- Maral Abdollahi (Ph.D., Hubbard School), current
- Smitha Muthya Sudheendra (Ph.D., CS&E), current
- Haiwei Ma (Ph.D., CS&E), *Understanding Expressive Writing in Online Health Communities*, current
- Ruyuan Wan (Masters, Linguistics), Plan A (thesis), *Riddikulus: Detection of Persuasion Techniques in Memes*, Masters Thesis, Passed
- Miguel Miguélez Díaz (Masters, Data Science), Plan B (capstone), *Contextual Saliency Variation of Sentiment Classifiers*, Masters Capstone, Passed
- Smitha Muthya Sudheendra (Masters, CS&E) *Computational Trust Approach to Measuring Trust in News Organizations*, Masters Thesis, Passed
- Yu Fang (Masters, CS&E), Directed Research, Passed
- Jared Grambihler (Undergraduate, CS&E), *Drug and Dosage Discovery for Opioid Use Disorder Recovery on Social Media Posts*, Undergraduate Honor Thesis, Passed
- David Ha (Undergraduate, CS&E), *Supplementing Bug Reports using Matched App Reviews*, Undergraduate Honor Thesis, Passed
- Risako Owan (Ph.D., CS&E), *A Computational Exploration of Modal Verbs: From Theories to Practice*, OPE/WPE, passed