Dongyeop Kang (DK)

412·736·4873 ◇ dongyeop@umn.edu ◇ dykang.github.io

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

Last updated on November 10, 2025

PROFESSIONAL EMPLOYMENT HISTORY

Assistant Professor University of Minnesota, Twin Cities, MN   2021 – present

Postdoctoral Scholar University of California, Berkeley, CA (*PI: Marti A. Hearst*)   2020 – 2021

Research Assistant Carnegie Mellon University (*PI: Eduard Hovy*)   2015 – 2020

Research Intern Facebook AI Research, Menlo Park, CA (*Mentors: Jason Weston*)   2018 Summer

Research Intern Allen Institute for Artificial Intelligence (AI2), Seattle, WA (*Mentors: Tushar Khot*)   2017 Summer

Research Intern Microsoft Research, Redmond, WA (*Mentors: Michael Gamon, Patrick Pantel*)   2016 Summer

Researcher Naver Labs and KAIST Institute, Korea (*Alternative Military Service*)   2012 – 2015

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

HONORS AND AWARDS

Google Research Gift Award	2025
Honorable Mention – VISxGenAI Workshop at IEEE VIS	2025
NSF NAIRR Pilot Award – National Science Foundation	2024
Best Student Paper Award Nomination – Interspeech	2024
Open Philanthropy Gift Award	2024
Grammarly Research Gift Award	2022-2024
3M Non-tenured Faculty Award (NTFA)	2022-2024
Cisco Research Award	2023
Best Paper Award – In2Writing Workshop at ACL	2022
Faculty Innovation Award – Sony Research	2022
Top Reviewer Recognition – NeurIPS and ACL	2018
Research Fellowship – Allen Institute for AI (AI2)	2018
Presidential Fellowship – CMU	2016
Startup Investment Prize – SMBA	2013

Ph.D. Scholarship – ILJU Foundation   2011
Excellent Performance Award – Microsoft Research Asia   2010
Summa Cum Laude – KAIST   2007

GRANTS AND RESEARCH GIFTS

Cognitive Foundation Models for Understanding User Cognitive States   2025
Google Research Gift
PI
Total Funds: $200,000

Interpreting Legal Reasoning Patterns in LLMs   2025
Thomson Reuters Research Gift
PI
Total Funds: $90,000

Scalable Reasoning and Efficient Model Adaptation   2025
Naver Corporation Research Gift
PI
Total Funds: $125,000

Personalized Phygital Brand Avatars for Small Businesses   2025
Digital Design Center (DDC) Computational Creativity Grant
PI (with Hyunjoo Im)
Total Funds: $35,000

Safe Path Planning for Autonomous Delivery Drones in Turbulent Urban Environments   2025
UMN MnRI Seed Grant
Co-PI (with Jiarong Hong, Changhyn Choi, Jianxun Wang)
Total Funds: $50,000

Scientific Foundational Model for Supporting Trustworthy Scholarly Writing   2025
NSF NAIRR Pilot
PI
Total Funds: NVIDIA DGX Cloud 280,000 GPU Hours

Improving Writing Assistants with Long-Term Coherence Modeling   2024
Grammarly Research Gift
PI
Total Funds: $150,000

Towards Expert Alignment: Collecting Legal Processing Benchmark from Lawyers   2024
Open Philanthropy — Lawyer Benchmark Collection
PI (with Daniel Schwarcz, Brett McDonnell)
Total Funds: $74,132

Aligning AI Models with Human Values: Formulation, Algorithms and Applications   2024
UMN Data Science Institute (DSI) Small Seed Grant
PI (with Mingyi Hong)
Total Funds: $15,000

Human-centric Natural Language Processing   2022 – 2024
3M NTFA Gift Awards
PI
Total Funds: $4,500

AI Assistant for Non-English Speaking Clinical Trial Participants   2023
UMN Clinical and Translational Science Institute (CTSI)
PI (with Jennifer Needle, Gwenyth Fischer)
Total Funds: $50,000

Computing Grants for Research   2023
OpenAI, Cohere AI, Google Cloud, Oracle Cloud, Amazon Web Services
PI
OpenAI: $25,000, Cohere AI: $5,000, Google Cloud: $6,000, Oracle Cloud: $1,000, AWS: $5,000

Active Instruction Tuning for Efficient and Collaborative Task Annotation   2023
Cisco Research
PI
Total Funds: $175,000

Writing Assistant for Professionals Using NLP + HCI Approaches   2023
Accenture Labs Gift
PI
Total Funds: $85,000

Collecting Human Explanations: Cognitively-inspired NLP Systems   2023
UMN Early Innovation Fund
PI
Total Funds: $10,431

MnRI Seed Grant Project   2023
UMN MnRI Seed Grant
Co-PI (with Karthik Desingh)
Total Funds: $40,000

Improving Writing Assistants with Planning and Iterative Revision   2022
Grammarly Research Gift
PI
Total Funds: $140,000

Goal-Oriented Decoding Control in Language Generation   2022
Sony Research
PI
Total Funds: $100,000

Setup Grant for Research Activities   2021
University of Minnesota

SERVICE AND PUBLIC OUTREACH

Program Chairs and Organizers

Co-organizer – AI for Future of Work (AIFoW) Workshop at CHI (under review)	2026
Co-organizer – Pluralistic Alignment Workshop at NeurIPS	2024
Area Chair – AAAI, EMNLP, ACL	2023-Present
Steering Committee – CustomNLP4U Workshop at EMNLP	2024
Advisory Board – Intelligent and Interactive Writing (In2Writing) Workshop at CHI	2024
Co-organizer – Intelligent and Interactive Writing (In2Writing) Workshop at CHI	2023
Co-organizer – Intelligent and Interactive Writing (In2Writing) Workshop at ACL	2022
Co-organizer – Minnesota NLP Seminar	2021-2022
Co-organizer – Controllable Generative Modeling CtrlGen Workshop at NeurIPS	2021

Program Committees (Conferences)

Conference on Human Factors in Computing Systems – CHI	2024-Present
Review Committee Member – ACL Rolling Review	2021-Present
Association for Computational Linguistics – ACL	2018-Present
Annual Conference of the NAACL – NAACL	2019, 2022
Conference on Empirical Methods in Natural Language Processing – EMNLP	2018-Present
Conference on Neural Information Processing Systems – NeurIPS	2018-Present
International Conference on Learning Representations – ICLR	2019-Present
International Conference on Machine Learning – ICML	2019

Program Committees (Workshops and Journals)

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

Department and University Service

Curriculum Committee – University of Minnesota	2023-2024
Faculty Recruiting Committee (Robotics/AI, HCC, ML areas) – University of Minnesota	2023-Present
Faculty Advisor, Undergraduate AI Club – University of Minnesota	2024-Present
Admissions Committee – University of Minnesota	2021-2023

PUBLICATIONS

Dissertations

D1 Linguistically Informed Language Generation: A Multifaceted Approach
Dongyeop Kang
PhD dissertation, Carnegie Mellon University 2020

Selected Publications

Refereed Conference Proceedings

C62 Align to Structure: Aligning Large Language Models with Structural Information
Zae Myung Kim, Anand Ramachandran, Farideh Tavazoee, Joo-Kyung Kim, Oleg Rokhlenko, and Dongyeop Kang
Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI) 2026

C61 Breaking Determinism: Stochastic Modeling for Reliable Off-Policy Evaluation in Ad Auctions
Hongseon Yeom, Jaeyoul Shin, Soojin Min, Jeongmin Yoon, Seunghak Yu, and Dongyeop Kang
Proceedings of International Conference on Web Search and Data Mining (WSDM) 2025
Oral

C60 How human is AI? A comparison of temporal meanings in 8 LLMs and 3 populations
Andreas Schramm, Karin de Langis, Jong Inn Park, Anh Thu Tong, Michael Mensink, Bin Hu, Khanh Chi Le, and Dongyeop Kang
Proceedings of American Association of Applied Linguistics (AAAL) 2025

C59 BBScoreV2: Learning Time-Evolution and Latent Alignment from Stochastic Representation
Tianhao Zhang, Zhecheng Sheng, Zhexiao Lin, Chen Jiang, and Dongyeop Kang
Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2025

C58 LawFlow : Collecting and Simulating Lawyers' Thought Processes
Debarati Das, Khanh Chi Le, Ritik Sachin Parkar, Karin De Langis, Brendan Madson, Chad M. Berryman, Robin M Willis, Daniel H Moses, Brett McDonnell, Daniel Schwarcz, and Dongyeop Kang
Conference on Language Modeling (COLM) 2025

C57 FlowForge: Guiding the Creation of Multi-agent Workflows with Interactive Visualizations as a Thinking Scaffold
Pan Hao, Dongyeop Kang, Nicholas Hinds, and Qianwen Wang
IEEE Transactions on Visualization and Computer Graphics (IEEE VIS) 2025

C56 How LLMs Comprehend Temporal Structure in Narratives: A Case Study in Cognitive Evaluation of LLMs
Karin de Langis, Jong Inn Park, Andreas Schramm, Bin Hu, Khanh Chi Le, and Dongyeop Kang
Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2025

C55 RoSTE: An Efficient Quantization-Aware Supervised Fine-Tuning Approach for Large Language Models
Quan Wei, Chung-Yiu Yau, Hoi To Wai, Yang Zhao, Dongyeop Kang, Youngsuk Park, and Mingyi Hong
International Conference on Machine Learning (ICML) 2025

C54 Joint Reward and Policy Learning with Demonstrations and Human Feedback Improves Alignment
Chenliang Li, Siliang Zeng, Zeyi Liao, Jiaxiang Li, Dongyeop Kang, Alfredo Garcia, and Mingyi Hong
International Conference on Learning Representations (ICLR) 2025
Spotlight (Top 5%)

C53 ISR-DPO: Aligning Large Multimodal Models for Videos by Iterative Self-Retrospective DPO
Daechul Ahn, Yura Choi, San Kim, Youngjae Yu, Dongyeop Kang, and Jonghyun Choi
Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI) 2025

C52 Chain-of-Instructions: Compositional Instruction Tuning on Large Language Models
Shirley Anugrah Hayati, Taehee Jung, Tristan Bodding-Long, Sudipta Kar, Abhinav Sethy, Joo-Kyung Kim, and Dongyeop Kang
Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI) 2025

C51 How Far Can We Extract Diverse Perspectives from Large Language Models?
Shirley Anugrah Hayati, Minhwa Lee, Dheeraj Rajagopal, and Dongyeop Kang
Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2024

C50 Dynamic Multi-Reward Weighting for Multi-Style Controllable Generation
Karin de Langis, Ryan Koo, and Dongyeop Kang
Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2024

C49 LearnerVoice: A Dataset of Non-Native English Learners' Spontaneous Speech
Haechan Kim, Junho Myung, Seoyoung Kim, Sungpah Lee, Dongyeop Kang, and Juho Kim
Conference of the International Speech Communication Association (Interspeech) 2024
Best Student Paper Nomination

C48 Talk Through It: End User Directed Manipulation Learning
Carl Winge, Adam Imdieke, Bahaa Aldeeb, Dongyeop Kang, and Karthik Desingh
IEEE Robotics and Automation Letters (RA-L) 2024

C47 Threads of Subtlety: Detecting Machine-Generated Texts Through Discourse Motifs
Zae Myung Kim, Kwang Hee Lee, Preston Zhu, Vipul Raheja, and Dongyeop Kang
Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2024

C46 Tuning Large Multimodal Models for Videos using Reinforcement Learning from AI Feedback
Daechul Ahn, Yura Choi, Youngjae Yu, Dongyeop Kang, and Jonghyun Choi
Annual Meeting of the Association for Computational Linguistics (ACL) 2024
Oral

C45 II-MMR: Identifying and Improving Multi-modal Multi-hop Reasoning in Visual Question Answering
Jihyung Kil, Farideh Tavazoee, Dongyeop Kang, and Joo-Kyung Kim
Annual Meeting of the Association for Computational Linguistics (ACL) Findings 2024

C44 Benchmarking Cognitive Biases in Large Language Models as Evaluators
Ryan Koo, Minhwa Lee, Vipul Raheja, Jong Inn Park, Zae Myung Kim, and Dongyeop Kang
Annual Meeting of the Association for Computational Linguistics (ACL) Findings 2024

C43 Consumer Engagement With AI-Powered Search Engines and Implications for the Future of Search Advertising
Gabriel Garlough-Shah, Jong Inn Park, Shirley Anugrah Hayati, Dongyeop Kang, and Jisu Huh
Association for Education in Journalism and Mass Communication (AEJMC) 2024

C42 Which Modality should I use - Text, Motif, or Image? : Understanding Graphs with Large Language Models
Debarati Das, Ishaan Gupta, Jaideep Srivastava, and Dongyeop Kang
Findings of the Association for Computational Linguistics: NAACL 2024 2024

C41 SkOTaPA: A Dataset for Skepticism Detection in Online Text after Persuasion Attempt
Smitha Muthya Sudheendra, Maral Abdollahi, Dongyeop Kang, Jisu Huh, and Jaideep Srivastava
Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) 2024

C40 BBScore: A Brownian Bridge Based Metric for Assessing Text Coherence
Zhecheng Sheng, Tianhao Zhang, Chen Jiang, and Dongyeop Kang
Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI) 2024

C39 Meta-Crafting: Improved Detection of Out-of-Distributed Texts via Crafting Metadata Space (Student Abstract)

Ryan Koo, Yekyung Kim, Dongyeop Kang, and Jaehyung Kim
Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI) Student Abstract 2024

C38 Dialogue Chain-of-Thought Distillation for Commonsense-aware Conversational Agents
Hyungjoo Chae, Yongho Song, Kai Tzu-iunn Ong, Taeyoon Kwon, Minjin Kim, Youngjae Yu, Dongha Lee, Dongyeop Kang, and Jinyoung Yeo
The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2023

C37 CoEdIt: State-of-the-art Text Editing by Task-Specific Instruction Tuning
Vipul Raheja, Dhruv Kumar, Ryan Koo, and Dongyeop Kang
The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP Findings) 2023
Media Coverage (MarTechPost) 10K+ download in HuggingFace as of 2024 May

C36 A Comparative Study on Textual Saliency of Styles from Eye Tracking, Annotations, and Language Models
Karin de Langis and Dongyeop Kang
Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL) 2023

C35 Quirk or Palmer: A Comparative Study of Modal Verb Frameworks with Annotated Datasets
Risako Owan, Maria Gini, and Dongyeop Kang
Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL) 2023

C34 Story Visualization by Online Text Augmentation with Context Memory
Daechul Ahn, Daneul Kim, Gwangmo Song, Seung Hwan Kim, Honglak Lee, Dongyeop Kang, and Jonghyun Choi
2023 IEEE/CVF International Conference on Computer Vision (ICCV) 2023

C33 Rebuilding Social Connection and Enhancing Advertising Effects Through the Nostalgic Appeal during the Pandemic
Katie Kim, Bugil Chang, Debarati Das, Smitha Sudheendra, Jisu Huh, Dongyeop Kang, and Jaideep Srivastava
Association for Education in Journalism and Mass Communication (AEJMC) 2023

C32 infoVerse: A Universal Framework for Dataset Characterization with Multidimensional Meta-information
Jaehyung Kim, Yekyung Kim, Karin de Langis, and Dongyeop Kang
Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL) 2023

C31 Rethinking Annotation: Can Language Learners Contribute?
Haneul Yoo, Rifki Afina Putri, Changyoon Lee, Youngin Lee, So-Yeon Ahn, Dongyeop Kang, and Alice Oh
2023

C30 Balancing Effect of Training Dataset Distribution of Multiple Styles for Multi-Style Text Transfer
Debarati Das, David Ma, and Dongyeop Kang
Findings of the Annual Meeting of the Association for Computational Linguistics (ACL) Findings 2023

C29 Prefer to Classify: Improving Text Classifiers via Auxiliary Preference Learning
Jaehyung Kim, Shin Jinwoo, and Dongyeop Kang
International Conference on Machine Learning (ICML) 2023

C28 Cluster-guided label generation in extreme multi-label classification
Taehee Jung, Joo-Kyung Kim, Sungjin Lee, and Dongyeop Kang
European Chapter of Association for Computational Linguistics (EACL) 2023

C27 StyLEx: Explaining Styles with Lexicon-Based Human Perception
Shirley Anugrah Hayati, Kyumin Park, Dheeraj Rajagopal, Lyle Ungar, and Dongyeop Kang
European Chapter of Association for Computational Linguistics (EACL) 2023

C26   Everyone's Voice Matters: Quantifying and Distinguishing Subjective Annotation Disagreement Using Demographic Information
      Ruyuan Wan, Jaehyung Kim, and Dongyeop Kang
      Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI) 2023
      Oral

C25   The Semantic Reader Project: Augmenting Scholarly Documents through AI-Powered Interactive Reading Interfaces
      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, and Daniel S. Weld
      Communications of ACM 2023

C24   Improving Iterative Text Revision by Learning Where to Edit from Other Revision Tasks
      Zae Myung Kim, Wanyu Du, Vipul Raheja, Dhruv Kumar, and Dongyeop Kang
      Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2022

C23   Understanding Iterative Revision from Human-Written Text
      Wanyu Du, Vipul Raheja, Dhruv Kumar, Zae Myung Kim, Melissa Lopez, and Dongyeop Kang
      Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 2022

C22   What Makes Better Augmentation Strategies? Augment Difficult but Not too Different
      Jaehyung Kim, Dongyeop Kang, Sungsoo Ahn, and Jinwoo Shin
      International Conference on Learning Representations (ICLR) 2022

C21   Does BERT Learn as Humans Perceive? Understanding Linguistic Styles through Lexica
      Shirley Anugrah Hayati, Dongyeop Kang, and Lyle Ungar
      Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP) *2021*

C20   Modeling Mathematical Notation Semantics in Academic Papers
      Hwiyeol Jo, Dongyeop Kang, Andrew Head, and Marti A. Hearst
      Findings on Empirical Methods in Natural Language Processing (EMNLP Findings) 2021

C19   Zero-shot Natural Language Video Localization
      Jinwoo Nam, Daechul Ahn, Dongyeop Kang, Seong Jong Ha, and Jonghyun Choi
      2021 IEEE/CVF International Conference on Computer Vision (ICCV) *2021*
      Oral (Top 3%)

C18   Style is NOT a single variable: Case Studies for Cross-Stylistic Language Understanding
      Dongyeop Kang and Eduard Hovy
      Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) 2021
      Oral

C17 Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols
Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst
Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI) 2021

C16 Plan ahead: Self-Supervised Text Planning for Paragraph Completion
Dongyeop Kang and Eduard Hovy
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020
Oral

C15 INSPIRED: Toward Sociable Recommendation Dialog Systems
Shirley A. Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020

C14 Posterior Calibrated Training on Sentence Classification Tasks
Taehee Jung, Dongyeop Kang, Hua Cheng, Lucas Mentch, and Thomas Schaaf
2020 Annual Conference of the Association for Computational Linguistics (ACL) 2020

C13 Earlier Isn't Always Better: Sub-aspect Analysis on Corpus and System Biases in Summarization
Dongyeop Kang*, Taehee Jung*, Lucas Mentch, and Eduard Hovy
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019

C12 Recommendation as a Communication Game: Self-Supervised Bot-Play for Goal-oriented Dialogue
Dongyeop Kang, Anusha Balakrishnan, Pararth Shah, Paul Crook, Y-Lan Boureau, and Jason Weston
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019

C11 (Male, Bachelor) and (Female, Ph.D) have different connotations: Parallelly Annotated Stylistic Language Dataset with Multiple Personas
Dongyeop Kang, Varun Gangal, and Eduard Hovy
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019
Oral

C10 Linguistic Versus Latent Relations for Modeling Coherent Flow in Paragraphs
Dongyeop Kang, Hiroaki Hayashi, Alan W Black, and Eduard Hovy
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2019

C9 Bridging Knowledge Gaps in Neural Entailment via Symbolic Models
Dongyeop Kang, Tushar Khot, Ashish Sabharwal, and Peter Clark
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2018

C8 AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples
Dongyeop Kang, Tushar Khot, Ashish Sabharwal, and Eduard Hovy
The 56th Annual Meeting of the Association for Computational Linguistics (ACL) 2018

C7 A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications
Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Madeleine van Zuylen, Sebastian Kohlmeier, Eduard Hovy, and Roy Schwartz
Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL) 2018

C6 Actionable email intent modeling with reparametrized RNN
Chu-Cheng Lin, Dongyeop Kang, Michael Gamon, Madian Khabsa, Ahmed Hassan Awadallah, and Patrick Pantel
Association for the Advancement of Artificial Intelligence (AAAI) 2018

C5 Detecting and Explaining Causes From Text For a Time Series Event
Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, and Eduard Hovy
Conference on Empirical Methods on Natural Language Processing (EMNLP) 2017

C4 Eventera: Real-time Event Recommendation System from Massive Heterogeneous Online Media
Dongyeop Kang, DongGyun Han, NaHea Park, Sangtae Kim, U Kang, and Soobin Lee
IEEE International Conference on Data Mining (ICDM) 2014

C3 Data/Feature Distributed Stochastic Coordinate Descent for Logistic Regression
Dongyeop Kang, Woosang Lim, Kijung Shin, Lee Sael, and U. Kang
Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management 2014

C2 Hetero-Labeled LDA: A Partially Supervised Topic Model with Heterogeneous Labels
Dongyeop Kang, Youngja Park, and Suresh Chari
ECML/PKDD 2014

C1 Multidimensional mining of large-scale search logs: a topic-concept cube approach
Dongyeop Kang, Daxin Jiang, Jian Pei, Zhen Liao, Xiaohui Sun, and Ho-Jin Choi
Web Search and Data Mining 2011

Refereed Workshop, Posters, Extended Abstracts, and Demos

W13 Joint Reward and Policy Learning with Demonstrations and Human Feedback Improves Alignment
Chenliang Li, Siliang Zeng, Zeyi Liao, Jiaxiang Li, Dongyeop Kang, Alfredo Garcia, and Mingyi Hong
International Conference on Learning Representations (ICLR) 2025

W12 Human-AI Collaborative Taxonomy Construction: A Case Study in Profession-Specific Writing Assistants
Minhwa Lee, Zae Myung Kim, Vivek A. Khetan, and Dongyeop Kang
CHI 2024 In2Writing Workshop 2024

W11 Confidence Calibration and Rationalization for LLMs via Multi-Agent Deliberation
Ruixin Yang, Dheeraj Rajagopal, Shirley Anugrah Hayati, Bin Hu, and Dongyeop Kang
ICLR 2024 Workshop on Reliable and Responsible Foundation Models 2024

W10 Annotation Imputation to Individualize Predictions: Initial Studies on Distribution Dynamics and Model Predictions
London Lowmanstone, Ruyuan Wan, Risako Owan, Jaehyung Kim, and Dongyeop Kang
Workshop on Perspectivist Approaches to NLP (NLPerspectives) @ECAI 2023 2023

W9 An Analysis of Reader Engagement in Literary Fiction through Eye Tracking and Linguistic Features
Rose Neis, Karin de Langis, Zae Myung Kim, and Dongyeop Kang
Workshop on Narrative Understanding (WNU) @ACL 2023 2023

W8 Decoding the End-to-end Writing Trajectory in Scholarly Manuscripts
Ryan Koo, Anna Martin, Linghe Wang, and Dongyeop Kang
Proceedings of the Second Workshop on Intelligent and Interactive Writing Assistants (In2Writing) at CHI 2023 https://minnesotanlp.github.io/REWARD_demo/, 2023

W7 Read, Revise, Repeat: A System Demonstration for Human-in-the-loop Iterative Text Revision
Wanyu Du, Zae Myung Kim, Vipul Raheja, Dhruv Kumar, and Dongyeop Kang
Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022) 2022
Best paper award

W6 User or Labor: An Interaction Framework for Human-Machine Relationships in NLP
Ruyuan Wan, Naome Etori, Karla Badillo-urquiola, and Dongyeop Kang
Proceedings of the Fourth Workshop on Data Science with Human-in-the-Loop (Language Advances) 2022

W5 Visualizing Cross-Lingual Discourse Relations in Multilingual TED Corpora
Zae Myung Kim, Vassilina Nikoulina, Dongyeop Kang, Didier Schwab, and Laurent Besacier
Proceedings of the 2nd Workshop on Computational Approaches to Discourse 2021

W4 Understanding Out-of-distribution: A Perspective of Data Dynamics
Dyah Adila and Dongyeop Kang
ICBINB@NeurIPS 2021

W3 GenAug: Data Augmentation for Finetuning Text Generators
Steven Y. Feng, Varun Gangal, Dongyeop Kang, Teruko Mitamura, and Eduard Hovy
Deep Learning Inside Out (DeeLIO) Workshop at EMNLP 2020 2020

W2 Document-Level Definition Detection in Scholarly Documents: Existing Models, Error Analyses, and Future Directions
Dongyeop Kang, Andrew Head, Risham Sidhu, Kyle Lo, Daniel Weld, and Marti A. Hearst
First Workshop on Scholarly Document Processing (SDP) at EMNLP 2020 2020

W1 News2Images: Automatically Summarizing News Articles into Image-Based Contents via Deep Learning
Jung-Woo Ha, Dongyeop Kang, Hyuna Pyo, and Jeonghee Kim
INRA@RecSys 2015

Archived Preprints and Papers Under Review

A17 Do LLMs Recognize Your Latent Preferences? A Benchmark for Latent Information Discovery in Personalized Interaction
Ioannis Tsaknakis, Bingqing Song, Shuyu Gan, Dongyeop Kang, Alfredo Garcia, Gaowen Liu, Charles Fleming, and Mingyi Hong
2025

A16 When Thoughts Meet Facts: Reusable Reasoning for Long-Context LMs
Soyeong Jeong, Taehee Jung, Sung Ju Hwang, Joo-Kyung Kim, and Dongyeop Kang
under review, 2025

A15 Are LLM Agents Behaviorally Coherent? Latent Profiles for Social Simulation
James Mooney, Josef Woldense, Zheng Robert Jia, Shirley Anugrah Hayati, My Ha Nguyen, Vipul Raheja, and Dongyeop Kang
2025

A14 Thinking Like A Lawyer In The Age Of Generative AI: Cognitive Limits On AI Adoption Among Lawyers
Daniel Schwarcz, Debarati Das, Dongyeop Kang, and Brett McDonnell
under review, 2025

A13 Toward Evaluative Thinking: Meta Policy Optimization with Evolving Reward Models
Zae Myung Kim, Chanoo Kim, Vipul Raheja, and Dongyeop Kang
under review, 2025

A12 A Framework for Robust Cognitive Evaluation of LLMs
Karin de Langis, Jong Inn Park, Bin Hu, Khanh Chi Le, Andreas Schramm, Michael C. Mensink, Andrew Elfenbein, and Dongyeop Kang
under review, 2025

A11 Stealing Creator's Workflow: A Creator-Inspired Agentic Framework with Iterative Feedback Loop for Improved Scientific Short-form Generation
Jong Inn Park, Maanas Taneja, Qianwen Wang, and Dongyeop Kang
2025

A10 Learning Explainable Dense Reward Shapes via Bayesian Optimization
Ryan Koo, Ian Yang, Vipul Raheja, Mingyi Hong, Kwang-Sung Jun, and Dongyeop Kang
under review, 2025

A9 ScholaWrite: A Dataset of End-to-End Scholarly Writing Process
Khanh Chi Le, Linghe Wang, Minhwa Lee, Ross Volkov, Luan Tuyen Chau, and Dongyeop Kang
under review, 2025

A8 CoTaxon: Enhancing Domain Expertise through Human-AI Collaborative Taxonomy Construction
Minhwa Lee, Zae Myung Kim, Vivek Khetan, Alex Kass, and Dongyeop Kang
under review, 2025

A7 LLM Leadership Coach for College Students: Exploring the Potential of LLM-Based Chatbots for College Leadership Coaching
Nga Do, Jessica Wang, Zach Roper, Dongyeop Kang, and Richard Landers
under review, 2025

A6 Anchors Aweigh! Sail for Optimal Unified Multi-Modal Representations
Minoh Jeong, Min Namgung, Zae Myung Kim, Dongyeop Kang, Yao-Yi Chiang, and Alfred Hero
under review, 2025

A5 SelectLLM: Can LLMs Select Important Instructions to Annotate?
Ritik Sachin Parkar, Jaehyung Kim, Jong Inn Park, and Dongyeop Kang
under review, 2024

A4 Shallow Synthesis of Knowledge in GPT-Generated Texts: A Case Study in Automatic Related Work Composition
Anna Martin-Boyle, Aahan Tyagi, Marti A. Hearst, and Dongyeop Kang
under review, 2024

A3 Under the Surface: Tracking the Artifactuality of LLM-Generated Data
Debarati Das, Karin De Langis, Anna Martin-Boyle, Jaehyung Kim, Minhwa Lee, Zae Myung Kim, Shirley Anugrah Hayati, Risako Owan, Bin Hu, Ritik Parkar, Ryan Koo, Jonginn Park, Aahan Tyagi, Libby Ferland, Sanjali Roy, Vincent Liu, and Dongyeop Kang
under review, 2024

A2 A Survey of Diffusion Models in Natural Language Processing
Hao Zou, Zae Myung Kim, and Dongyeop Kang
under review, 2023

A1 Complex Mathematical Symbol Definition Structures: A Dataset and Model for Coordination Resolution in Definition Extraction
Anna Martin-Boyle, Andrew Head, Kyle Lo, Risham Sidhu, Marti A. Hearst, and Dongyeop Kang
under review, 2023

TALKS

Invited Talks and Guest Lectures

Invited speaker – May Brodbeck Outreach series, UMN Department of Philosophy   Oct 2025

Invited speaker – Seoul National University	Aug 2025
Invited speaker – Korea University	Aug 2025
Invited speaker – Allen Institute for AI (AI2)	Jun 2025
Invited speaker – The Undergraduate AI Club	Sep 2024
Invited speaker – Megagon Labs	May 2024
Guest lecture – UMN JOUR8500 Computational Communication Research	Mar 2024
Invited speaker – Thomson Reuters Labs	Nov 2023
Invited speaker – Pennsylvania State University (PSU)	Sep 2023
Invited speaker – UMN NLP/IE group	Sep 2023
Invited speaker – 3M NTFA Symposium	Aug 2023
Invited speaker – Microsoft Research	May 2023
Invited panelist – UMN Law School	May 2023
Interview – Echo (student-run news site of St. Louis Park High School)	Apr 2023
Invited speaker – UMN Korean-American Scientists and Engineers Association (KSEA)	Mar 2023
Invited speaker – Google People+AI Research (PAIR)	Feb 2023
Invited speaker – Hyundai AI	Dec 2022
Invited speaker – NSF ROSE-HUB	Nov 2022
Invited speaker – 3M NTFA Symposium	Sep 2022
Invited speaker – Yonsei University	Jul 2022
Invited speaker – KAIST CS	Jul 2022
Invited speaker – LG AI Research	Jul 2022
Invited speaker – Samsung Research	Jul 2022
Invited speaker – Grammarly	Jul 2022
Invited speaker – Thomson Reuters Labs	Jul 2022
Invited speaker – USC ISI AI Seminar	Mar 2022
Guest lecture – UMN The AI Undergraduate Club	Mar 2022
Invited speaker – Naver AI US	Feb 2022
Invited speaker and panelist – CMU Language Technology Institute (LTI)	Feb 2022
Invited speaker – GaTech NLP Seminar	Dec 2021
Invited speaker – Minnesota Robotic Institute Colloquium	Nov 2021
Invited speaker – KAIST SE Colloquium	Nov 2021
Invited speaker – Naver Labs Europe	Nov 2021
Invited speaker – Grammarly AI	Aug 2021
Invited speaker – SNU Summer AI School	Aug 2021
Invited speaker – University of Wisconsin - Madison, CS Department	Mar 2021
Invited speaker – University of Rochester, CS Department	Mar 2021
Invited speaker – University of Minnesota - Twin Cities, CS&E Colloquium	Mar 2021
Invited speaker – KAIST AI Department	Feb 2021
Invited speaker – KAIST Colloquium	Nov 2020
Invited speaker – GIST EECS Colloquium	Nov 2020
Invited speaker – Berkeley AI Research (BAIR) Workshop	Jul 2020
Invited speaker – CMU LTI Seminar	Jul 2020
Invited speaker – POSTECH AI Seminar	Jul 2020

TEACHING AND CURRICULUM DEVELOPMENT

Courses

CSCI 8980 - Large Language Model - Spring 2026 (Scheduled)
CSCI 5541 - Natural Language Processing - Spring 2023, Fall 2023, Spring 2024, Fall 2024, Fall 2025
CSCI 5980 - Natural Language Processing (NLP) with Deep Learning - Fall 2022
CSCI 8980 - Introduction to Natural Language Processing (NLP) Research - Spring 2022

ADVISING AND MENTORING

Current Ph.D. Students

Karin de Langis, DDF Fellow
Risako Owan (co-advised with Gini)
Zae Myung Kim, 3M & DDF Fellow
Shirley Anugrah Hayati
James Mooney, Amazon MLSys Fellow
Shuyu Gan (co-advised with Hong)

Past Ph.D Students

Debarati Das (co-advised with Srivastava), Ph.D. UMN CS 2025 → Senior Applied Scientist at Microsoft
Jaehyung Kim, Visiting Ph.D. 2023 → CMU Postdoc → Assistant Professor, Yonsei AI

Current Masters and Undergraduate Students

Ruizi Wang
William Walker
Chi Le
Rimika Dhara
Ryan Peters

Past Masters and Undergraduate Students

Ryan Koo, CS&E Undergraduate/Masters → Amazon
Ritik Parkar, Robotics Masters → Thomson Reuters
Linghe Wang, CS&E Undergraduate/Masters → Complexity AI
Jianing Wen, CS&E Undergraduate/Masters → PhD @Northeastern
Bin Hu, CS&E Undergraduate/Masters → Masters @UMaryland
Preston Zhu, CS&E Undergraduate
Vincent Liu, CS&E Undergraduate
Ross Volkov, CS&E Undergraduate
Robert Jia, CS&E Undergraduate → Transferred to GaTech
Tripp Dow, CS&E Undergraduate
Minhwa Lee, visiting Master student (UMass) → Mayo Clinic
Ishaan Gupta, CS&E Undergraduate
Ruixin Yang, visiting Undergraduate (UBC) → GaTech Masters

Jong Inn Park, Data Science Masters, Plan B → Research Assistant, UMN

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 → EMBL-EBI

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

Petros Karypis, CS&E Undergraduate → UCSD CS Masters → UCSD CS PhD

Wanyu Du, Grammarly intern → University of Virginia → Amazon AWS

Ph.D./Masters/Undergraduate Committee Member

Zhongxing Zhang (Ph.D., CS&E), *BiMind: A Behavior-aware Dual-Head Reasoning Model for Incorrect Information Detection*, Ph.D. Oral Prelim Exam (OPE), 2026

Nga Do (Ph.D., Industrial Organizational Psychology), Ph.D. Oral Prelim Exam (OPE), Passed, 2026

Basel Hussein (Ph.D., Learning Technologies, Education & Human Development), *Now What? An Empirical Description of Goal-Setting, Misalignment, and Learning Along the Way in a Digital Puzzle-Based Video Game*, Ph.D. thesis defense, Passed, 2024

Jiyoon Pyo (Ph.D., CS&E), *Leveraging Large Language Models for Generating Labeled Mineral Site Record Linkage Data*, OPE/WPE, Passed, 2024

Zae Myung Kim (Ph.D., CS&E), OPE/WPE, Passed, 2024

Shirley Anugrah Hayati (Ph.D., CS&E), OPE/WPE, Passed, 2024

Anna Martin (Ph.D., CS&E), OPE/WPE, Passed, 2024

London Lowmanstone (Ph.D., CS&E), OPE/WPE, Passed, 2024

Tianhao Zhang (Ph.D., Applied Mathematics), *Model of the astrocytic morphology remodeling process*, Ph.D. thesis proposal, Passed, 2023

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