

Sherry Tongshuang Wu

HCII, School of Computer Science
Carnegie Mellon University
Newell-Simon Hall 3525
Pittsburgh, PA 15213

+1 206-910-2784
sherryw@cs.cmu.edu
[@sherryw](http://cs.cmu.edu/~sherryw)
[@tongshuangwu](https://twitter.com/tongshuangwu)

ACADEMIC EXPERIENCE

- 2022– Carnegie Mellon University, Daniel P. Siewiorek Assistant Professor of HCII
Human-Computer Interaction Institute (CMU HCII)
Language Technology Institute (CMU LTI)
- 2016–22 University of Washington, Research Assistant
with Jeffrey Heer, Dan Weld
Pitfalls in *status quo* human-AI interactions.
Principles and tools for enhanced NLP model analysis.
Controllable generators for model analysis and improvement.

EDUCATION

- 2016–22 Ph.D. in Computer Science and Engineering
University of Washington, Seattle, WA
Thesis: Interactive AI Model Debugging and Correction
Advisor: Jeffrey Heer, Dan Weld
Committee: Marco Tulio Ribeiro, Noah Smith, Mari Ostendorf
- 2016–18 M.S. in Computer Science and Engineering
University of Washington, Seattle, WA
- 2012–16 B.Eng. in Computer Science and Engineering
Hong Kong University of Science and Technology, Hong Kong, Hong Kong
Advisor: Huamin Qu
- 2014 Exchange student in Computer Science and Engineering
University of Michigan, Ann Arbor, MI

INDUSTRY EXPERIENCE

- 2021 Google Brain/PAIR, Research Intern & Part-time Student Researcher
with Carrie Cai, Michael Terry
Transparent & controllable human-AI collaborations via multi-step problem-solving.
- 2019 Microsoft Research, Research Intern
with Marco Tulio Ribeiro
Behavioral testing for NLP models covering broad model capabilities.
- 2018–19 Apple Inc., Full-time Intern & Part-time Intern
with Chris DuBois, Kayur Patel, Kanit Wongsuphasawat, Donghao Ren, Charlie Maalouf
Structural analysis for unstructured text datasets.
- 2017 Microsoft Research, Research Intern
with Bongshin Lee, Ece Kamar, Saleema Amersh
Uncertainty-aware data labeling and visual refinement.
- 2015 Microsoft Research Asia, Research Intern
with Weiwei Cui
De-cluttering statistical graphs.

SELECTED HONORS AND AWARDS

- 2025 NeurIPS 2025 Responsible Foundation Model Workshop Best Paper Award
- 2026 AAAI 2026 FrontierIR Workshop Best Paper Award
- 2025 Gemini Academic Research Award
- 2024 Google Academic Research Award

- 2024 Amazon Research Awards
- 2024 AIED 2024 Best Paper Award
- 2024 AIED 2024 Honorable Mention Award
- 2024 AIED 2024 Best Interactive Event Award
- 2023 CSCW 2023 Best Demo Award
- 2023 IUI 2023 Honorable Mention Award
- 2022 CHI 2022 Honorable Mention Award
- 2020 Rising Stars in EECS Workshop (UC Berkeley)
A highly selective workshop based on academic excellence and commitment to advancing equity and inclusion.
- 2020 ACL 2020 Best Paper Award
- 2016–17 Faithful Steward Endowed Fellowship in Computer Science & Engineering
- 2012–16 Scholarship Scheme for Continuing Undergraduate Students
- 2016 IEEE PacificVis 2016 Honorable Mention Award
- 2016 IEEE PacificVis 2016 Best Notes Paper

PUBLICATIONS

** denotes equal contribution*

Manuscripts and Pre-prints

- 2026 P.1 Aditya Mittal, Ryan Shar, Zichu Wu, Shyam Agarwal, Tongshuang Wu, Chris Donahue, Ameet Talwalkar, Wayne Chi, Valerie Chen. **Comparing Developer and LLM Biases in Code Evaluation.** ArXiv 2026
- P.2 Canwen Wang, Angela Chen, Catherine Bao, Siwei Jin, Yee Kit Chan, Jessica R Mindel, Sijia Xie, Holly Swartz, Tongshuang Wu, Robert E Kraut, Haiyi Zhu. **Modeling Multi-Party Interaction in Couples Therapy: A Multi-Agent Simulation Approach.** ArXiv 2026
- P.3 Xuhui Zhou, Weiwei Sun, Qianou Ma, Yiqing Xie, Jiarui Liu, Weihua Du, Sean Welleck, Yiming Yang, Graham Neubig, Sherry Tongshuang Wu, Maarten Sap. **Mind the Sim2Real Gap in User Simulation for Agentic Tasks.** ArXiv 2026
- 2025 P.4 Rulin Shao, Akari Asai, Shannon Zejiang Shen, Hamish Ivison, Varsha Kishore, Jingming Zhuo, Xinran Zhao, Molly Park, Samuel G Finlayson, David Sontag, Tyler Murray, Sewon Min, Pradeep Dasigi, Luca Soldaini, Faeze Brahma, Wen-tau Yih, Tongshuang Wu, Luke Zettlemoyer, Yoon Kim, Hannaneh Hajishirzi, Pang Wei Koh. **DR Tulu: Reinforcement Learning with Evolving Rubrics for Deep Research.** ArXiv 2025
- P.5 Yixiao Zeng, Tianyu Cao, Danqing Wang, Xinran Zhao, Zimeng Qiu, Morteza Ziyadi, Tongshuang Wu, Lei Li. **RARE: Retrieval-Aware Robustness Evaluation for Retrieval-Augmented Generation Systems.** ArXiv 2025

Peer-reviewed Journal and Conference Publications

- 2026 P.6 Keyu He, Qianou Ma, Wayne Chi, Valerie Chen, Tongshuang Wu. **RECAP: An End-to-End Platform for Capturing, Replaying, and Analyzing AI-Assisted Programming Interactions.** ACL Demo Track 2026
- P.7 Shannon Zejiang Shen, Valerie Chen, Ken Gu, Alexis Ross, Zixian Ma, Jillian Ross, Alex Gu, Chenglei Si, Wayne Chi, Andi Peng, Jocelyn J Shen, Ameet Talwalkar, David Sontag, Tongshuang Wu. **Scaling Collaborative Effort with Agents.** ACL Findings 2026
- P.8 Chenyang Yang, Yike Shi, Qianou Ma, Michael Xieyang Liu, Christian Kästner, Tongshuang Wu. **What Prompts Don't Say: Understanding and Managing Underspecification in LLM Prompts.** ACL Findings 2026
- P.9 Qianou Ma, Megan Chai, Yike Tan, Jihun Choi, Jini Kim, Erik Harpstead, Geoff Kauffman, Tongshuang Wu. **"GenAI Defaults to Bias!" Gamify AI literacy through Reflections on Prompts.** AIED 2026
- P.10 Eason Chen, Xinyi Tang, Yvonne Zhao, Meiyi Chen, Meryam Elmir, Elizabeth McLaughlin, Mingyu Yuan, Yumo Wang, Shyam Agarwal, Jared Cochrane, Jionghao Lin, Tongshuang Wu, Ken Koedinger. **Practice Less, Explain More: LLM-Supported Self-Explanation Improves Explanation Quality on Transfer Problems in Calculus.** AIED 2026
- P.11 Shreya Bali, Riku Arakawa, Peace Odiase, Tongshuang Wu, Mayank Goel. **Evidotes: Integrating Scientific Evidence and Anecdotes to Support Uncertainties Triggered by Peer Health Posts.** CHI 2026
- P.12 Chang Liu, Qinyi Zhou, Xinjie Shen, Xingyu Bruce Liu, Tongshuang Wu, Xiang 'Anthony' Chen. **Behavioral Indicators of Overreliance During Interaction with Conversational Language Models.** CHI 2026
- P.13 Qianou Ma, Kenneth Koedinger, Tongshuang Wu. **Not Everyone Wins with LLMs: Behavioral Patterns and Pedagogical Implications in AI-assisted Data Analysis.** CHI 2026

- P.14 Xinran Zhao, Aakanksha Naik, Jay DeYoung, Joseph Chee Chang, Jena D. Hwang, Tongshuang Wu, Varsha Kishore. **Improving Attributed Long-form Question Answering with Intent Awareness**. ICLR 2026
- P.15 Fengyu Cai, Tong Chen, Xinran Zhao, Sihao Chen, Hongming Zhang, Sherry Tongshuang Wu, Iryna Gurevych, Heinz Koepl. **Revela: Dense Retriever Learning via Language Modeling**. ICLR 2026
- P.16 Lexin Zhou, Lorenzo Pacchiardi, Fernando Martínez-Plumed, Katherine M. Collins, Yael Moros-Daval, Seraphina Zhang, Qinlin Zhao, Yitian Huang, Luning Sun, Jonathan E. Prunty, Zongqian Li, Pablo Sánchez-García, Kexin Jiang Chen, Pablo A. M. Casares, Jiyun Zu, John Burden, Behzad Mehrbakhsh, David Stillwell, Manuel Cebrian, Jindong Wang, Peter Henderson, Sherry Tongshuang Wu, Patrick C. Kyllonen, Lucy Cheke, Xing Xie, José Hernández-Orallo. **General Scales Unlock AI Evaluation with Explanatory and Predictive Power**. Nature 2026
- P.17 Chentianye Xu, Jionghao Lin, Tongshuang Wu, Vincent Alevan, Kenneth R. Koedinger. **Improving Automated Feedback Systems for Tutor Training in Low-Resource Scenarios through Data Augmentation**. TLT 2026
- 2025 P.18 Shijie Xia, Xuefeng Li, Yixin Liu, Tongshuang Wu, Pengfei Liu. **Evaluating Mathematical Reasoning Beyond Accuracy**. AAAI 2025
- P.19 Qianou Ma*, Dora Zhao*, Xinran Zhao, Chenglei Si, Chenyang Yang, Ryan Louie, Ehud Reiter, Diyi Yang+, Tongshuang Wu+. **SPHERE: An Evaluation Card for Human-AI Systems**. ACL Findings 2025
- P.20 Tongshuang Wu, Haiyi Zhu, Maya Albayrak, Alexis Axon, Amanda Bertsch, Wenxing Deng, Ziqi Ding, Bill Guo, Sireesh Gururaja, Tzu-Sheng Kuo, Jenny T Liang, Ryan Liu, Ihita Mandal, Jeremiah Milbauer, Xiaolin Ni, Namrata Padmanabhan, Subhashini Ramkumar, Alexis Sudjianto, Jordan Taylor, Ying-Jui Tseng, Patricia Vaidos, Zhijin Wu, Wei Wu, Chenyang Yang. **LLMs as Workers in Human-Computational Algorithms? Replicating Crowdsourcing Pipelines with LLMs**. CHI Case Study 2025
- P.21 Jushaan Singh Kalra, Xinran Zhao, To Eun Kim, Fengyu Cai, Fernando Diaz, Tongshuang Wu. **MoR: Better Handling Diverse Queries with a Mixture of Sparse, Dense, and Human Retrievers**. EMNLP 2025
- P.22 Yilin Zhang, Xinran Zhao, Zora Zhiruo Wang, Chenyang Yang, Jiayi Wei, Tongshuang Wu. **cAST: Enhancing Code Retrieval-Augmented Generation with Structural Chunking via Abstract Syntax Tree**. EMNLP Findings 2025
- P.23 Chenyang Yang, Tesi Xiao, Michael Shavlovsky, Christian Kästner, Tongshuang Wu. **Orbit: A Framework for Designing and Evaluating Multi-objective Rankers**. IUI 2025
- P.24 Xuhui Zhou, Zhe Su, Sophie Feng, Jiayu Zhou, Jen-tse Huang, Hsien-Te Kao, Spencer Lynch, Svitlana Volkova, Tongshuang Wu, Anita Woolley, Hao Zhu, Maarten Sap. **SOTOPIA-S4: a user-friendly system for flexible, customizable, and large-scale social simulation**. NAACL Demo Track 2025
- P.25 Vijay Viswanathan, Yanchao Sun, Shuang Ma, Xiang Kong, Meng Cao, Graham Neubig, Tongshuang Wu. **Checklists Are Better Than Reward Models For Aligning Language Models**. NeurIPS Spotlight 2025
- P.26 Qianou Ma, Weirui Peng, Chenyang Yang, Hua Shen, Kenneth Koedinger, Tongshuang Wu. **What Should We Engineer in Prompts? Training Humans in Requirement-Driven LLM Use**. TOCHI 2025
- 2024 P.27 Saumya Gandhi, Ritu Gala, Vijay Viswanathan, Tongshuang Wu, Graham Neubig. **Better Synthetic Data by Retrieving and Transforming Existing Datasets**. ACL Findings 2024
- P.28 Xinran Zhao, Hongming Zhang, Xiaoman Pan, Wenlin Yao, Dong Yu, Tongshuang Wu, Jianshu Chen. **Fact-and-Reflection (FaR) Improves Confidence Calibration of Large Language Models**. ACL Findings 2024
- P.29 Qiaomu Ma, Hua Shen, Kenneth Koedinger, Tongshuang Wu. **How to Teach Programming in the AI Era? Using LLMs as a Teachable Agent for Debugging**. AIED 2024 Best Paper, Best Interactive Event
- P.30 Atharva Naik, Jessica Ruhan Yin, Anusha Kamath, Qianou Ma, Sherry Tongshuang Wu, Charles Murray, Majd Sakr, Carolyn P. Rose. **Generating Situated Reflection Triggers About Alternative Solution Paths: A Case Study in Generative AI for Computer-Supported Collaborative Learning**. AIED 2024 Best Paper Nominee
- P.31 Chenyang Yang, Yining Hong, Grace A. Lewis, Tongshuang Wu, Christian Kästner. **What Is Wrong with My Model? Identifying Systematic Problems with Semantic Data Slicing**. ASE 2024
- P.32 Atharva Naik, Jessica Ruhan Yin, Anusha Kamath, Qianou Ma, Sherry Tongshuang Wu, R. Charles Murray, Christopher Bogart, Majd Sakr, Carolyn P. Rose. **Providing Tailored Reflection Instructions in Collaborative Learning Using Large Language Models**. BERA 2024
- P.33 Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, Brad A. Myers. **Selenite: Scaffolding Online Sensemaking with Comprehensive Overviews Elicited from Large Language Models**. CHI 2024
- P.34 Tzu-Sheng Kuo, Aaron Halfaker, Zirui Cheng, Jiwoo Kim, Meng-Hsin Wu, Tongshuang Wu, Ken Holstein, Haiyi Zhu. **Wikibench: Community-Driven Data Curation for AI Evaluation on Wikipedia**. CHI 2024
- P.35 Cheng Qian, Xinran Zhao, Tongshuang Wu. **"Merge Conflicts!" Exploring the Impacts of External Distractors to Parametric Knowledge Graphs**. CoLM 2024
- P.36 Xinran Zhao, Tong Chen, Sihao Chen, Hongming Zhang, Tongshuang Wu. **Beyond Relevance: Evaluate and Improve Retrievers on Perspective Awareness**. CoLM 2024

- P.37 Chenyang Zhao, Xueying Jia, Vijay Viswanathan, Graham Neubig, Tongshuang Wu. **Self-Guide: Better Task-Specific Instruction Following via Self-Synthetic Finetuning**. CoLM 2024
- P.38 Yujia Qin, Shengding Hu, Yankai Lin, Weize Chen, Ning Ding, Ganqu Cui, Zheni Zeng, Yufei Huang, Chaojun Xiao, Chi Han, Yi Ren Fung, Yusheng Su, Huadong Wang, Cheng Qian, Runchu Tian, Kunlun Zhu, Shihao Liang, Xingyu Shen, Bokai Xu, Zhen Zhang, Yining Ye, Bowen Li, Ziwei Tang5, Jing Yi, Yuzhang Zhu, Zhenning Dai, Lan Yan, Xin Cong, Yaxi Lu, Weilin Zhao, Yuxiang Huang, Junxi Yan, Xu Han, Xian Sun, Dahai Li, Jason Phang, Cheng Yang, Tongshuang Wu, Heng Ji, Zhiyuan Liu, Maosong Sun. **Tool Learning with Foundation Models**. Computing Surveys 2024
- P.39 Ian Wu, Sravan Jayanthi, Vijay Viswanathan, Simon Rosenberg, Sina Pakazad, Tongshuang Wu, Graham Neubig. **Synthetic Multimodal Question Generation**. EMNLP Findings 2024
- P.40 Chenglei Si, Navita Goyal, Tongshuang Wu, Chen Zhao, Shi Feng, Hal Daumé III, Jordan Boyd-Graber. **Large Language Models Help Humans Verify Truthfulness – Except When They are Convincingly Wrong**. NAACL 2024
- P.41 Shayne Longpre, Robert Mahari, Anthony Chen, Naana Obeng-Marnu, Damien Sileo, William Brannon, Niklas Muennighoff, Nathan Khazam, Jad Kabbara, Kartik Perisetla, Xinyi (Alexis) Wu, Enrico Shippole, Kurt Bollacker, Tongshuang Wu, Luis Villa, Sandy Pentland, Deb Roy, Sara Hooker. **A Large Scale Audit of Dataset Licensing and Attribution in AI**. Nature Machine Intelligence 2024
- P.42 Lindia Tjuatja, Valerie Chen, Tongshuang Wu, Ameet Talwalkar, Graham Neubig. **Do LLMs Exhibit Human-Like Response Biases? A Case Study in Survey Design**. TACL 2024
- 2023 P.43 Vijay Viswanathan, Luyu Gao, Tongshuang Wu, Pengfei Liu, Graham Neubig. **DataFinder: Scientific Dataset Recommendation from Natural Language Descriptions**. ACL 2023
- P.44 Logan Stapleton, Jordan Taylor, Sarah Fox, Tongshuang Wu, Haiyi Zhu. **Seeing Seeds Beyond Weeds: Green Teaming Generative AI for Beneficial Uses**. ArXiv 2023
- P.45 Yiming Zhang, Sravani Nanduri, Liwei Jiang, Tongshuang Wu, Maarten Sap. **BiasX: "Thinking Slow" in Toxic Content Moderation with Explanations of Implied Social Biases**. EMNLP 2023
- P.46 Jeremiah Milbauer, Ziqi Ding, Zhijin Wu, Tongshuang Wu. **From Nuisance to News Sense: Augmenting the News with Cross-document Evidence and Context**. EMNLP Demo Track 2023
- P.47 Vijay Viswanathan, Chenyang Zhao, Amanda Bertsch, Tongshuang Wu, Graham Neubig. **Prompt2Model: Generating Deployable Models from Natural Language Instructions**. EMNLP Demo Track 2023
- P.48 Chenyang Yang, Rishabh Rustogi, Rachel Brower-Sinning, Grace Lewis, Christian Kaestner, Tongshuang Wu. **Beyond Testers' Biases: Guiding Model Testing with Knowledge Bases using LLMs**. EMNLP Findings 2023
- P.49 Tongshuang Wu, Hua Shen, Jeffrey Heer, Daniel S. Weld, Marco Tulio Ribeiro. **ScatterShot: Interactive In-context Example Curation for Text Transformation**. IUI 2023 Honorable Mention
- P.50 Kaustubh D Dhole, Varun Gangal, Sebastian Gehrmann, Aadesh Gupta, Zhenhao Li, Saad Mahamood, Abinaya Mahendiran, Simon Mille, Ashish Srivastava, Samson Tan, Tongshuang Wu, Jascha Sohl-Dickstein, Jinho D Choi, Eduard Hovy, Ondrej Dusek, Sebastian Ruder, et al.. **NL-Augmenter: A Framework for Task-Sensitive Natural Language Augmentation**. NEJLT 2023
- P.51 Vijay Viswanathan, Kiril Gashteovski, Carolin Lawrence, Tongshuang Wu, Graham Neubig. **Large Language Models Enable Few-Shot Clustering**. TACL 2023
- P.52 Patrick Fernandes, Aman Madaan, Emmy Liu, António Farinhas, Pedro Henrique Martins, Amanda Bertsch, José G. C. de Souza, Shuyan Zhou, Tongshuang Wu, Graham Neubig, André F. T. Martins. **Bridging the Gap: A Survey on Integrating (Human) Feedback for Natural Language Generation**. TACL 2023
- P.53 Hyeonsu Kang, Tongshuang Wu, Joseph Chee Chang, Aniket Kittur. **Synergi: A Mixed-Initiative System for Scholarly Synthesis and Sensemaking**. UIST 2023
- 2022 P.54 Bingsheng Yao, Dakuo Wang, Tongshuang Wu, Toby Jia-Jun Li, Mo Yu, Ying Xu. **It is AI's Turn to Ask Humans a Question: Question and Answer Pair Generation for Children Storybooks with FairytaleQA Dataset**. ACL 2022
- P.55 Ying Xu, Dakuo Wang, Mo Yu, Daniel Ritchie, Bingsheng Yao, Tongshuang Wu, Zheng Zheng, Toby Jia-Jun Li, Nora Bradford, Branda Sun, Tran Bao Hoang, Yisi Sang, Yufang Hou, Xiaojuan Ma, Diyi Yang, Nanyun Peng, Zhou Yu, Mark Warschauer. **Fantastic Questions and Where to Find Them: FairytaleQA -- An Authentic Dataset for Narrative Comprehension**. ACL 2022
- P.56 Tongshuang Wu*, Alexis Ross*, Hao Peng, Matthew E. Peters, Matt Gardner. **Tailor: Generating and Perturbing Text with Semantic Controls**. ACL 2022
- P.57 Hua Shen, Tongshuang Wu, Wenbo Guo, Ting-Hao 'Kenneth' Huang. **Are Shortest Rationales the Best Explanations for Human Understanding?.** ACL 2022
- P.58 Tongshuang Wu, Michael Terry, Carrie J. Cai. **AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts**. CHI 2022
- P.59 Zheng Zhang, Ying Xu, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, Toby Jia-Jun Li.

StoryBuddy: A Human-AI Collaborative Agent for Parent-Child Interactive Storytelling with Flexible Parent Involvement. CHI 2022

- P.60 Jiao Sun, Tongshuang Wu, Yue Jiang, Ronil Awalegaonkar, Xi Victoria Lin, Diyi Yang. **Pretty Princess vs. Successful Leader: Gender Roles in Greeting Card Messages**. CHI 2022 **Honorable Mention**
- P.61 Tongshuang Wu*, Ellen Jiang*, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, Carrie J. Cai. **PromptChainer: Chaining Large Language Model Prompts through Visual Programming**. CHI LBW 2022
- P.62 Yun Wang, Zhitao Hou, Leixian Shen, Tongshuang Wu, Jiaqi Wang, He Huang, Haidong Zhang, Dongmei Zhang. **Towards Natural Language-Based Visualization Authoring**. TVCG 2022
- 2021 P.63 Tongshuang Wu, Marco Tulio Ribeiro, Jeffrey Heer, Daniel S. Weld. **Polyjuice: Generating Counterfactuals for Explaining, Evaluating, and Improving Models**. ACL 2021
- P.64 Tongshuang Wu*, Gagan Bansal*, Joyce Zhou+, Raymond Fok+, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, Daniel S. Weld. **Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance**. CHI 2021
- P.65 Xingbo Wang, Yao Ming, Tongshuang Wu, Haipeng Zeng, Yong Wang, Huamin Qu. **DeHumor: Visual Analytics for Decomposing Humor**. TVCG 2021
- 2020 P.66 Marco Tulio Ribeiro, Tongshuang Wu, Carlos Guestrin, Sameer Singh. **Beyond Accuracy: Behavioral Testing of NLP Models with CheckList**. ACL 2020 **Best Paper**
- P.67 Alison Smith-Renner, Ron Fan, Melissa Birchfield, Tongshuang Wu, Jordan Boyd-Graber, Daniel S. Weld, Leah Findlater. **No Explainability without Accountability: An Empirical Study of Explanations and Feedback in Interactive ML**. CHI 2020
- P.68 Tongshuang Wu, Kanit (Ham) Wongsuphasawat, Donghao Ren, Kayur Patel, Chris DuBois. **Tempura: Query Analysis with Structural Templates**. CHI 2020
- P.69 Tongshuang Wu*, Zhihang Dong*, Sicheng Song, Mingrui Zhang. **Interactive Attention Model Explorer for Natural Language Processing Tasks with Unbalanced Data Sizes**. PacificVis 2020
- 2019 P.70 Tongshuang Wu, Marco Tulio Ribeiro, Jeffrey Heer, Daniel S. Weld. **Errudite: Scalable, Reproducible, and Testable Error Analysis**. ACL 2019
- P.71 Yang Shi, Maoran Xu, Rongwen Zhao, Hao Fu, Tongshuang Wu, Nan Cao. **Interactive Context-Aware Anomaly Detection Guided by User Feedback**. THMS 2019
- P.72 Tongshuang Wu, Daniel S. Weld, Jeffrey Heer. **Local Decision Pitfalls in Interactive Machine Learning: An Investigation into Feature Selection in Sentiment Analysis**. TOCHI 2019
- 2016 P.73 Yun Wang, Tongshuang Wu, Zhutian Chen, Huamin Qu, Qiong Luo. **STAC: Enhancing Stacked Graphs for Time Series Analysis**. PacificVis 2016
- P.74 Tongshuang Wu, Yuan Yao, Yuqing Duan, Xinzhi Fan, Huamin Qu. **NetworkSeer: Visual Analysis for Social Network in MOOCs**. PacificVis 2016 **Best Paper**
- P.75 Tongshuang Wu, Yingcai Wu, Conglei Shi, Huamin Qu, Weiwei Cui. **PieceStack: Toward Better Understanding of Stacked Graphs**. TVCG 2016 **Honorable Mention**
- P.76 Qiaomu Shen, Tongshuang Wu, Haiyan Yang, Yanhong Wu, Huamin Qu, Weiwei Cui. **NameClarifier: A Visual Analytics System for Author Name Disambiguation**. TVCG 2016

Posters, Extended Abstracts, Workshop Papers and Technical Reports

- 2025 W.1 Xinran Zhao, Boyuan Zheng, Chenglei Si, Haofei Yu, Ken Ziyu Liu, Runlong Zhou, Ruochen Li, Tong Chen, Xiang Li, Yiming Zhang, Tongshuang Wu. **The Ramon Llull's Thinking Machine for Automated Ideation**. COLM LM4Sci 2025
- 2024 W.2 Zirui Wang, Xinran Zhao, Simon Stepputtis, Woojun Kim, Tongshuang Wu, Katia Sycara, Yaqi Xie. **HiMemFormer: Hierarchical Memory-Aware Transformer for Multi-Agent Action Anticipation**. Video-Language Models Workshop @ NeurIPS 2024
- 2023 W.3 Chenyang Yang, Rachel Brower-Sinning, Grace A. Lewis, Christian Kästner, Tongshuang Wu. **Capabilities for Better ML Engineering**. AAAI SafeAI 2023
- W.4 Yuanchen Bai, Raoyi Huang, Vijay Viswanathan, Tzu-Sheng Kuo, Tongshuang Wu. **Measuring Adversarial Datasets**. AACL ART of Safety 2023
- W.5 Qianou Christina Ma, Tongshuang Wu, Kenneth Koedinger. **Is AI the Better Programming Partner? Human-Human Pair Programming vs. Human-AI pAIr Programming**. AIED2023 Empowering Education with LLMs 2023
- W.6 Hua Shen, Tongshuang Wu. **Parachute: Evaluating Interactive Human-LM Co-writing Systems**. CHI In2Writing 2023
- W.7 Hua Shen, Chieh-Yang Huang, Tongshuang Wu, Ting-Hao (Kenneth) Huang. **ConvXAI: Delivering Heterogeneous AI Explanations via Conversations to Support Human-AI Scientific Writing**. CSCW Demo Track 2023 **Best Demo**
- 2022 W.8 Zheng Zheng, Ying Xu, Yanhao Wang, Tongshuang Wu, Bingsheng Yao, Daniel Ritchie, Mo Yu, Dakuo Wang, Toby

- Jia-Jun Li. **Building a Storytelling conversational Agent through Parent-AI Collaboration**. AAAI AI4ED 2022
- 2021 W.9 Tongshuang Wu. **Principles and Interactive Tools for Evaluating and Improving the Behavior of Natural Language Processing models**. CHI DC 2021
- 2018 W.10 Halden Lin, Tongshuang Wu, Kanit (Ham) Wongsuphasawat, Yejin Choi, Jeffrey Heer. **Visualizing Attention in Sequence-to-Sequence Summarization Models**. VAST 2018

Patent

- 2023 PT.1 Carrie Cai, Tongshuang Wu, Michael Terry. **Transparent and Controllable Human-AI Interaction via Chaining of Machine-Learned Language Models**. US Patent US 2023/0112921 A1 2023
- 2022 PT.2 Ajit Narayanan, Subhashini Venugopalan, Tongshuang Wu, Shanqing Cai, Michael Terry, Meredith Morris, Carrie Cai. **Providing Suggestions of Expanded Text from Abbreviated Text Input**. (Defensive Publication) 2022

GRANTS

- G.1 **Accountable Collaboration: Attributing Contributions in Human-AI Workflows (2026-27)**
PI, \$50,000, from CMU Block Center
- G.2 **Multimodal Agents to Aid Exploratory Data Analysis (2026-27)**
Co-PI, \$240,000, from Jane Street, with Daniel Fried
- G.3 **Digital Twins that Capture and Apply Expert Tacit Knowledge (2026-27)**
PI, \$120,000, from CMU-Teleperformance Research Center, with Niki Kittur
- G.4 **Gemini Academic Program Award (2026-27)**
PI, \$50,000 (compute credits), from Google Research
- G.5 **Identifying and Improving Higher-Order Thinking in College Math (2026-27)**
PI, \$140,000, from CMU Learning Engineering Institute, with Ken Koedinger
- G.6 **Evaluations and Benchmarks (2026-27)**
Co-PI, \$75,000, from Google DeepMind, with Ken Koedinger
- G.7 **Capturing and Applying Expert Tacit Knowledge (2025-26)**
PI, \$100,000, from Bosch, with Niki Kittur
- G.8 **Requirements-Driven AI App Synthesis and Testing (2025-26)**
PI, \$100,000, from Apple
- G.9 **Trajectory Retrieval for Compositional Reasoning in LLM Agents (2025-26)**
PI, \$100,000, from Cisco Research
- G.10 **PFI-TT: Thinking and Talking Books (2024-25)**
Co-PI, \$550,000, from NSF, with Umut Acar
- G.11 **SCH: Training Mental Health Supporters with Virtual Patients and Automated Feedback (2024-28)**
Co-PI, \$900,000, from NIMH, with Haiyi Zhu, Holly Swartz, Robert Kraut, Diyi Yang
- G.12 **Contextualized Domain-Aware Dynamic RAG-LLM Evaluation Ecosystem for Interactive Data (2024-25)**
Co-PI, \$750,000, from Amazon, with Mona Diab, Fernando Diaz, Lei Li, Sherry Wu, Chenyan Xiong, Akari Asai
- G.13 **Research and tools for Human-AI collaboration (2024-25)**
PI, \$60,000, from Google Gift
- G.14 **Explicitly Train Humans towards Human-AI Collaboration (2024-25)**
PI, \$100,000, from Google Gift
- G.15 **Generating Deployable Models from Natural Language Instructions through Adaptive Data Curation (2024-25)**
PI, \$120,000 (50k compute credits + 70k cash), from Amazon
- G.16 **OpenAI API Access Program (2024-25)**
PI, \$10,000 (compute credits), from Amazon
- G.17 **GCP Credit Award: Gemma Academic Program (2024-25)**
PI, \$5,000 (compute credits), from Google
- G.18 **AgInteract: LLM-based Human-AI Teaming Simulation Environment (2023-25)**
Co-PI, \$997,979, from DARPA, with Maarten Sap, Carolyn Rose, Daniel Fried, Anita Woolley
- G.19 **Robust Collaborative AI to Enable Effective Human Decision Making under Uncertainty and Time-Stress (2023-27)**
Co-PI, \$1,960,873, from Office of Naval Research, with Katia Sycara
- G.20 **Quantifying the Characteristics of LLMs Evaluation (2023-24)**
PI, \$10,000, from Google Research
- G.21 **From General-purpose AI to Use Cases (2023-24)**

PI, \$20,000, from Adobe Research

G.22 Enhance CS Education with Teachable Agents (2023-24)

PI, \$20,000 (compute credits), from Oracle Cloud

G.23 Broadening Large Language Models' Impact with Small Language Models (2023-27)

Co-PI, \$20,000 (compute credits), from Microsoft Research, with Graham Neubig

TEACHING EXPERIENCE

Instructor

- 2026 05-317/617 Design of AI Products and Services (CMU)
- 2025 05-898B Mini Data Science for Product Managers (CMU)
05-898A Mini HCI for Product Managers (CMU)
05-499/899 Human-Centered NLP (LLM) (CMU)
- 2024 17-445/645/745 Machine Learning in Production (CMU)
- 2023-24 05-391/891 Designing Human Centered Software (CMU)
- 2023 05-499/899 Human-Centered NLP (CMU)

Guest Lecture

- 2025 Human-AI Collaboration: How AIs Augment Human Teammates (William & Mary)
Making AI Collaborators Work for Imperfect Humans (UC Berkeley)
- 2024 Human Interactions with Code Gen Models (Carnegie Mellon University)
Interacting with Large Language Models (Carnegie Mellon University)
- 2023 Human-Centered AI (University of South California)
- 2022 Visualization and Machine Learning (Carnegie Mellon University)
Interacting with Large Language Models (Carnegie Mellon University)
Designing Human-Centered, AI-Infused Software (Carnegie Mellon University)
Visualizing Text Summarization Models (Carnegie Mellon University)
- 2021 HCI+AI Interaction (University of Notre Dame)
- 2019 Model Interpretability (University of Washington)

Conference Tutorial

- 2025 ACL 2025: How AIs Augment Human Teammates
- 2024 NAACL 2024: Human-AI Interaction in the Age of LLMs Models
- 2023 EMNLP 2023: Designing, Evaluating, and Learning from Humans Interacting with NLP Models

Teaching Assistant

- 2019 CSE 512 Data Visualization (University of Washington)
- 2018 CSE 442 Data Visualization (University of Washington)

MENTORING EXPERIENCE

Advisees

- PhD Vijay Viswanathan (CMU LTI, *co-advisor: Graham Neubig*). Synthesize Supervision Signals
Christina Ma (CMU HCII, *co-advisor: Ken Koedinger*). Train People for Future of Work
Chenyang Yang (CMU S3D, *co-advisor: Christian Kästner*). Human-Centered ML Engineering
Xinran Zhao (CMU LTI). Information Seeking for Complex Tasks
Jessie Mindel (CMU HCII). Simulated Agents and Collective Sensemaking
Zheyuan Zhang (CMU LTI). Human-Agent Interaction
Eunsu Kim (CMU LTI). AI Contributions to Human Tasks.
- Master Yiyang (Diana) Wang (CMU HCII). End-User Prompt Disambiguation. Now PhD student at Georgia Tech.
Yuanchen (Sophie) Bai (CMU Heinz). NLP dataset characterization
Raoyi (Cathy) Huang (CMU Heinz). NLP dataset characterization. Now PhD student at Cornell.
Atharva Naik (CMU LTI). LLM in CS education. Now PhD student at CMU.
Jushaan Kalra (CMU MIIS). Multi-domain Retrieval. Now software engineer at Snowflake.
Keyu He (CMU MIIS). Eval of Human Prompt Ability

Callum Zhao (CMU MS in EE). *Build Effective Agentic Skills*
 Yilin Zhang (CMU MIIS). *Code Retrieval with AST. Now software engineer at Google.*

Visit Cheng Qian (Tsinghua University). *LLM hallucination. Now PhD student at UIUC.*
 Yuan Tian (ZJU). *AI for Data Analysis*

Undergrad Alex Cheung (CMU IS). *LLM sensemaking copilot*
 Samridhhi Bhardwaj (CMU CS). *LLM sensemaking copilot*
 Alina Chen *LLM sensemaking copilot*
 Yashika Batra (CMU CS). *LLM sensemaking copilot*
 Shaan Lehal (CMU CS). *LLM sensemaking copilot*
 Cassandra Shi (CMU CS). *Requirement-driven LLMs. Now PhD at NYU-Shanghai.*

Thesis Committee

PhD Jordan Taylor (CMU). *Queer Computing: Rethinking Users, Uses and Utility Functions*
 Annalisa Szymanski (University of Notre Dame). *Incorporating Domain Expertise Into Large Language Model Evaluation Workflows for Complex Tasks*
 Sireesh Gururaja (CMU). *NLP in the Last Mile: Characterizing and Resisting Incentives Towards Generality*
 Jessica Huynh (CMU). *Automatic Personalized Evaluation*
 Brihi Joshi (USC). *Towards Useful and Richer User Signals for Personalization*
 Angela Chen (CMU). *Empirically Grounded LLM-based Virtual Patients for Psychotherapy Training*
 Will Epperson (CMU). *Interactive Data Profiling Systems for Data Programming*
 Steven Moore (CMU). *Creating and Evaluating Pedagogically Valid Assessments at Scale*
 Yoonjoo Lee (KAIST). *Aligning AI Agents with How Humans Understand Knowledge*
 Hyeonsu Kang (CMU). *Accelerating Innovation through AI-Powered Conceptual Abstraction and Interaction Design*
 Kundan Krishna (CMU). *Improving the reliability of summarization models*
 Hua Shen (Penn State). *Towards Useful AI Interpretability via Interactive AI Explanations*
 Jason Wu (CMU). *Computational Understanding of User Interfaces*

Master Shreya Bali (CMU). *Tools to facilitate working on Machine Learning in the Industry*
 Ihita Mandal (CMU). *Accessible Descriptions for Surprising Clusters in Scatterplots*

Prior to CMU

PhD Yi Guo (Tongji University). *Co-supervised with Nan Cao. Natural-language-based visualization generation.*
 Sebastin Santy (UW). *The design and creation of an HCI+NLP research playbook.*
 Jiao Sun (USC). *Co-supervised with Diyi Yang. Gender bias in NLP datasets.*

Master Joyce Zhou (UW; Now at Cornell). *Co-supervised with Dan Weld & Gagan Bansal. Human-AI teaming.*
 Halden Lin (UW; Now at Apple Inc.). *Attention visualization for NLP models.*
 Akshat Shrivastava (UW; Now at Meta). *Active learning for sequence labeling.*

PROFESSIONAL SERVICE

Organizing Committees

2026 From Human-Human Collaboration to Human-Agent Collaboration: A Vision, Design Philosophy, and an Empirical Framework for Achieving Successful Partnerships Between Humans and LLM Agents (CHI 2026)

2025 MMU-RAG: the Massive Multi-Modal User-Centric Retrieval-Augmented Generation Benchmark (NeurIPS 2025 Competition)

2025 Tutorial: How AIs Augment Human Teammates (ACL 2025)

2025 BiAlign: Bidirectional Human-AI Alignment (ICLR 2025 (Workshop) & CHI 2025 (SIG))

2024 TREW: Workshop on Trust and Reliance in Evolving Human-AI Workflows (CHI 2024)

2024 Tutorial: Human-AI Interaction in the Age of LLMs (NAACL 2024)

2023 Tutorial: Designing, Evaluating, and Learning from Humans Interacting with NLP Models (EMNLP 2023)

2022 SSSL: Sharing Stories and Lessons Learned Workshop (EMNLP 2022)

2022 TRAIT: Workshop on Trust and Reliance in AI-Human Teams (CHI 2022-23)

2022 NL-Augmenter (part of GEM: Workshop for Generation, Evaluation, Metrics, ACL 2021)

Program Committees

AI AAAI 2022, AAAI HCOMP 2022-23, ACM FAccT 2022, NeurIPS XAI 2021

NLP ACL 2023/25, EMNLP 2023-25, NAACL 2025, COLM 2024

HCI CHI 2023-24/26, ACM IUI 2022-23, IUI TExSS 2022, CHI HCXAI 2021

Paper Reviewing

AI ICLR 2026, NeurIPS 2025, Nature 2025, NeurIPS 2022, AAAI 2022, AKBC 2021, ACM Computing Surveys 2021
HCI ACM CHI 2019-22, TOCHI 2021/25, UIST 2018/20/22, IUI 2020, CSCW 2020, TiiS 2022
Special recognition for outstanding reviews ACM CHI, IUI
NLP ACL 2020/25/26, EACL 2021, NAACL 2021
Viz. IEEE VIS 2017-23, TVCG 2021, EuroVis 2021, PacificVis 2018/20, ChinaVis 2017-19

Community Service

2024-26 Committee member, CMU Faculty Senate
Represented the HCII department in the CMU Faculty Senate.

2024-25 Committee member, CMU K&L Gates Award Selection Committee
Selected awardees who have inspired their fellow students to love learning through a combination of intellect, high scholarly achievement, engagement with others, and character.

2024 Reviewer, Department of Energy Office Proposal Panel

2023-25 Committee member, AAAI/SIGAI Doctoral Dissertation Award
Selected candidates for AAAI and ACM SIGAI thesis award.

2023-24 Committee member, CMU HCII PhD Admission Committee
Committee member, CMU HCII Undergraduate Admission Committee
Leader, Postdoc Mentoring Group
Led a bi-weekly mentorship group for Postdocs within PhD HCII.

2023 Reviewer, NSF Proposal Panel

2022-25 Committee member, ACM/SCS Thesis Nomination & Award Committee
Selected awardees for CMU Dissertation Award, as well as candidates for ACM Thesis Award.

2021 Co-organizer, UW Allen School Women's Research Day
An outreach event to women and nonbinary people in research.
Co-organizer, UW Allen School Pre-Application Mentoring Service (PAMS)
A program supporting 107 potential CS PhD applicants, with 80% from underrepresented communities.
Coordinator, UW Allen School Diverse Genders in Research Events
Course design mentor, UW AVELA (A Vision for Electronic Literacy & Access)
Mentored undergraduate students to develop curriculum for high-school web development courses.

2020 Student volunteer, IEEE VIS 2020
Student contributor, UW Allen School Strategic Plan for Diversity, Equity & Inclusion
Subcommittee Student Assistant, ACM SIGCHI 2021

2019 Reviewer, UW Allen School Graduate Admission Committee

2018 Co-leader, UW Interactive Systems Seminar

2013 Community tutor, HKUST Connect

MEDIA COVERAGE

2026 [CMU Tool Prevents Anxiety Spirals When Searching for Medical Advice Online](#)
CMU SCS News, 2025.5

2025 [5 Ways to Stay Smart When Using Gen AI, Explained by Computer Science Professors](#)
CNET, 2025.3
[Carnegie Mellon Community Shines at SXSW 2025 — an Intersection of Culture, Tech and Innovation](#)
CMU News, 2025.3
[You're a Computer Science Major. Don't Panic.](#)
New York Times, 2025.11

2024 [SCS Faculty Receive Google Academic Research Awards](#)
CMU School of Computer Science News, 2024.1

2023 [Debugging Imperfect AI](#)
CMU User Experience Association (UXA) Newsletter, 2023.2

- CMU & Tsinghua U's Prompt2Model Generates Deployable Models Following Natural Language Instructions
Synced Review, 2023.8
- Researchers from CMU and Tsinghua University Propose Prompt2Model: A General Purpose Method that Generates Deployable AI Models from Natural Language Instructions
MarkTechPost, 2023.8
- AI Researchers Uncover Ethical, Legal Risks to Using Popular Data Sets
The Washington Post, 2023.1
- 15% of datasets for fine-tuning language models use Wikipedia
Wikimedia, 2023.11
- Bigger isn't always better when it comes to large language models
Axios, 2023.12
- MIT, Cohere for AI, others launch platform to track and filter audited AI datasets
VentureBeat, 2023.1
- 2020 Allen School researchers earn Best Paper Award at ACL 2020
Allen School News, 2020.8
- AI researchers create testing tool to find bugs in NLP from Amazon, Google, and Microsoft
VentureBeat, 2020.7
- How Should We Do Error Analysis? A Lesson for NLP developers [Chinese]
AI Technology Review, 2020.8
- 2019 Experimental results and error reporting, Ethics and NLP, Distillation
Distillation vol. 2., SemEval 2020

INVITED TALKS

- 2026 AI in Action: Industry Leaders and Education's Path Forward
TREC 2026: ReWiring Learning, invited panelist (2026.2)
- 2025 How to be a Smarter AI User
SXSW 2025 (2025.3)
- Practical AI Systems: From General-Purpose AI to (the Right) Specific Use Cases
Peking University, Wangxuan Institute of Computer Technology (2023.12)
HEAL: Human-centered Evaluation and Auditing of Language Models @ CHI 2024 (2024.5)
HCI+NLP Workshop @ NAACL 2024 (2024.5)
MIT NLP Seminar (2024.9)
Learning Machine Seminar Series (LMSS) @ Cornell Tech (2024.1)
CXBT NLP-CoP Distinguished Speaker Series (2025.1)
- Making AI Collaborators Work for Imperfect Humans
The Societal AI Lecture Series @ MSR Asia (2025.11)
Schmidt Sciences Human Futures in the Age of AI Symposium (2025.11)
AWS Agentic AI (2025.11)
- Human-Agent Interaction: The Process Matters, Too
CMU Agent Workshop 2025 (2025.4)
- 2024 How do LLMs Change the Practical Impact of Explanations?
Natural Language Reasoning and Structured Explanations Workshop @ ACL 2024 (2024.8)
- 2023 Education and the Future of Work
CMU Generative AI Innovation Incubator, invited panelist (2023.6)
- Practical AI Systems and Effective Human-AI collaboration
NEC Labs Europe (2023.8)
HCI@KAIST Colloquium (2023.1)
- LLMs and the Infrastructure of CSCW
CSCW 2023 panelist (2023.1)
- 2022 **Advanced NLP: Prompting**
Peking University, Wangxuan Institute of Computer Technology (2022.3)
- Interactive AI Model Debugging and Correction
Carnegie Mellon University, Ameet Group Meeting (2022.9)

Carnegie Mellon University, LTI Colloquium (2022.9)
University of Washington, DUB Seminar (2022.7)
University of Washington, Paul G. Allen School of CSE (2022.6)
MIT Computer Science & Artificial Intelligence Laboratory (2022.3)
Stanford University, Computer Science Department (2022.3)
Princeton University, Computer Science Department (2022.3)
Cornell University, Computer Science Department (2022.3)
Carnegie Mellon University, Human-Computer Interaction Institute (2022.3)
Peking University, Wangxuan Institute of Computer Technology (2022.3)
UT Austin, Computer Science Department (2022.2)
University of Chicago, Data Science Institute (2022.2)
Hong Kong University of Science and Technology, Computer Science Department (2022.2)
PennState, Young Achievers Symposium (2022.1)

Model-in-the-loop Data Collection: What Roles does the Model Play?

DADC: Dynamic Adversarial Data Collection, NAACL 2022 (2022.7)

2021 Generating and Perturbing Text with Semantic Controls

Allen Institute for Artificial Intelligence, All-AI2 Meeting (2021.8)

Machines in the Loop: Explainability, Transparency, and Rich Interaction

ACL InterNLP 2021, invited panelist (2021.8)

Transparent and Controllable Collaboration with Large Language Models

Google Film Sprint: Fluid Language Integrating Muse (2021.1)

Google PAIR: People+AI Research Initiative (2021.7)

Principles and Tools for Evaluating and Improving NLP Models

Google Research (2021.7)

ACM CHI Doctoral Consortium (2021.5)

NLP with Friends (2021.4)

Interactive Data Exploration System (IDEAS) lab, Shandong University (2021.4)

Human+AI: the Relationships, the Goals, the Challenges

University of Notre Dame, Human-Centered Computing Research (2021.1)

Hong Kong University of Science and Technology, VisLab (2021.1)

2020 Behavioral Testing of NLP Models

AI Technology Review (2020.8)

AI Time PhD (2020.7)

UCLA, Center for Vision, Cognition, Learning and Autonomy (VCLA) (2020.7)

2019 Scalable, Reproducible, and Testable Error Analysis

UW CSE 512, as guest lecturer (2019.5)

Allen Institute for Artificial Intelligence, All-AI2 Meeting (2019.5)

Apple Inc., Knowledge Graph Team Seminar (2019.3)

Robust AI Event: Research & Reality (2019.5)