

Yizhong Wang

Key Center
601 108th Ave NE, 11F
Bellevue, WA 98004

Website: yizhongw.github.io
Email: yizhong01@gmail.com
Phone: +1 (949)-241-0294

Education

09/2019–06/2025	University of Washington , Seattle, USA Ph.D. in Computer Science Advisors: Hannaneh Hajishirzi, Noah A. Smith Thesis: Scalable Data Paradigms for Steering General-Purpose Language Models
09/2016–06/2019	Peking University , Beijing, China M.Sc. in Computer Science Advisor: Sujian Li
09/2012–06/2016	Shanghai Jiao Tong University , Shanghai, China B.Sc. in Computer Science and Technology

Employment

06/2025–Present	ByteDance Seed, Bellevue, USA Research Scientist
2026 (forthcoming)	The University of Texas at Austin, Austin, USA Assistant Professor , Computer Science Department
01/2023–03/2025	Allen Institute for Artificial Intelligence, Seattle, USA Student Researcher with Daniel Khashabi, Pradeep Dasigi
09/2021–12/2021	Facebook (Meta) AI Research, London, UK Research Intern with Sebastian Riedel
12/2020–05/2021	Allen Institute for Artificial Intelligence, Seattle, USA Research Intern with Tushar Khot, Ashish Sabharwal
09/2016–07/2019	Institute of Computational Linguistics, Beijing, China Research Assistant with Sujian Li
10/2018–02/2019	Allen Institute for Artificial Intelligence, Seattle, WA, USA Research Intern with Matt Gardner, Sameer Singh
05/2018–09/2018	Microsoft Research Asia, Beijing, China Research Intern with Furu Wei
06/2017–02/2018	Baidu NLP, Beijing, China Research Intern with Kai Liu, Yajuan Lyu

Awards & Honors

2024	Best Theme Paper of ACL 2024 (for OLMo)
2024	GeekWire Innovation of the Year 2024 (for OLMo)
2020	Best Paper Award of CCL 2020 (for LiveQA)
2017	Outstanding Paper Award of ACL 2017 (for Staged Discourse Parsing)
2017–2018	Founder Scholarship, Peking University
2014–2016	Chun-Tsung Scholarship (by Nobel laureate Tsung-Dao Lee), Shanghai Jiao Tong University
2013–2015	Academic Excellence Fellowship, Shanghai Jiao Tong University

Publications

Google Scholar: <https://scholar.google.com/citations?user=y5zpqdAAAAAJ>

Semantic Scholar: <https://www.semanticscholar.org/author/Yizhong-Wang/1705260>

* denotes equal contribution

Peer-Reviewed Conference/Journal Papers

- [C31] Nathan Lambert, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, Lester James V. Miranda, Alisa Liu, Nouha Dziri, Xinxu Lyu, Yuling Gu, Saumya Malik, Victoria Graf, Jena D. Hwang, Jiangjiang Yang, Ronan Le Bras, Oyvind Tafjord, Chris Wilhelm, Luca Soldaini, Noah A. Smith, **Yizhong Wang***, Pradeep Dasigi, and Hannaneh Hajishirzi. “Tülu 3: Pushing Frontiers in Open Language Model Post-Training”. In: *Proceedings of COLM*. 2025.
- [C30] Zhiyuan Zeng, **Yizhong Wang**, Hannaneh Hajishirzi, and Pang Wei Koh. “EvalTree: Profiling Language Model Weaknesses via Hierarchical Capability Trees”. In: *Proceedings of COLM*. 2025.
- [C29] Lj Miranda*, **Yizhong Wang***, Yanai Elazar, Sachin Kumar, Valentina Pyatkin, Faeze Brahman, Noah A. Smith, Hannaneh Hajishirzi, and Pradeep Dasigi. “Hybrid Preferences: Learning to Route Instances for Human vs. AI Feedback”. In: *Proceedings of ACL*. 2025.
- [C28] Seungone Kim, Juyoung Suk, Xiang Yue, Vijay Viswanathan, Seongyun Lee, **Yizhong Wang**, Kiril Gashteovski, Carolin Lawrence, Sean Welleck, and Graham Neubig. “Evaluating language models as synthetic data generators”. In: 2025.
- [C27] Wenhao Wu, **Yizhong Wang**, Guangxuan Xiao, Hao Peng, and Yao Fu. “Retrieval Head Mechanistically Explains Long-Context Factuality”. In: *Proceedings of ICLR*. 2025. **(Oral, top 1.8%)**.
- [C26] Kevin Xu, Yeganeh Kordi, Kate Sanders, **Yizhong Wang**, Adam Byerly, Jack Zhang, Benjamin Van Durme, and Daniel Khashabi. “Tur[k]ingBench: A Challenge Benchmark for Web Agents”. In: *Proceedings of NAACL*. 2025.
- [C25] Hamish Ivison, **Yizhong Wang**, Jiacheng Liu, Zeqiu Wu, Valentina Pyatkin, Nathan Lambert, Noah A. Smith, Yejin Choi, and Hannaneh Hajishirzi. “Unpacking DPO and PPO: Disentangling Best Practices for Learning from Preference Feedback”. In: *Proceedings of NeurIPS*. 2024.

- [C24] Alisa Liu, Xiaochuang Han, **Yizhong Wang**, Yulia Tsvetkov, Yejin Choi, and Noah A. Smith. “Tuning Language Models by Proxy”. In: *Proceedings of COLM*. 2024. (*Spotlight, top 7%*).
- [C23] Abhika Mishra, Akari Asai, Vidhisha Balachandran, **Yizhong Wang**, Graham Neubig, Yulia Tsvetkov, and Hannaneh Hajishirzi. “Fine-Grained Hallucination Detection and Editing for Language Models”. In: *Proceedings of COLM*. 2024.
- [C22] Dirk Groeneveld, Iz Beltagy, Pete Walsh, Akshita Bhagia, Rodney Kinney, Oyvind Tafjord, Ananya Harsh Jha, Hamish Ivison, Ian Magnusson, **Yizhong Wang**, Shane Arora, David Atkinson, Russell Authur, Khyathi Raghavi Chandu, Arman Cohan, Jennifer Dumas, Yanai Elazar, Yuling Gu, Jack Hessel, Tushar Khot, William Merrill, Jacob Morrison, Niklas Muennighoff, Aakanksha Naik, Crystal Nam, et al. “OLMo: Accelerating the Science of Language Models”. In: *Proceedings of ACL*. 2024. (*Best Theme Paper Award*).
- [C21] Bowen Zhao*, Zander Brumbaugh*, **Yizhong Wang***, Hannaneh Hajishirzi, and Noah A. Smith. “Set the Clock: Temporal Alignment of Pretrained Language Models”. In: *Proceedings of ACL Findings*. 2024.
- [C20] Akari Asai, Zeqiu Wu, **Yizhong Wang**, Avirup Sil, and Hannaneh Hajishirzi. “Self-RAG: Learning to Retrieve, Generate, and Critique Through Self-Reflection”. In: *Proceedings of ICLR*. 2024. (*Oral, top 1.2%*).
- [C19] Qingqing Cao, Sewon Min, **Yizhong Wang**, and Hannaneh Hajishirzi. “BTR: Binary Token Representations for Efficient Retrieval-Augmented Language Models”. In: *Proceedings of ICLR*. 2024.
- [C18] **Yizhong Wang***, Hamish Ivison*, Pradeep Dasigi, Jack Hessel, Tushar Khot, Khyathi Raghavi Chandu, David Wadden, Kelsey MacMillan, Noah A. Smith, Iz Beltagy, et al. “How Far Can Camels Go? Exploring the State of Instruction Tuning on Open Resources”. In: *Proceedings of NeurIPS*. 2023. (*Spotlight, top 3%*).
- [C17] Yushi Hu, Benlin Liu, Jungo Kasai, **Yizhong Wang**, Mari Ostendorf, Ranjay Krishna, and Noah A. Smith. “TIFA: Accurate and Interpretable Text-to-Image Faithfulness Evaluation with Question Answering”. In: *Proceedings of ICCV*. 2023.
- [C16] **Yizhong Wang**, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. “Self-Instruct: Aligning Language Models with Self-Generated Instructions”. In: *Proceedings of ACL*. 2023. (*ACL 2023 Most Influential Paper #1 by Paper Digest*).
- [C15] Hongjin Su, Weijia Shi, Jungo Kasai, **Yizhong Wang**, Yushi Hu, Mari Ostendorf, Wen-tau Yih, Noah A. Smith, Luke Zettlemoyer, and Tao Yu. “One Embedder, Any Task: Instruction-Finetuned Text Embeddings”. In: *Proceedings of ACL Findings*. 2023.
- [C14] Hamish Ivison, Akshita Bhagia, **Yizhong Wang**, Hannaneh Hajishirzi, and Matthew Peters. “HINT: Hypernetwork Instruction Tuning for Efficient Zero- and Few-Shot Generalisation”. In: *Proceedings of ACL*. 2023.
- [C13] **Yizhong Wang***, Swaroop Mishra*, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, et al. “Super-NaturalInstructions: Generalization via Declarative Instructions on 1600+ NLP Tasks”. In: *Proceedings of EMNLP*. 2022. (*EMNLP 2022 Most Influential Paper #2 by Paper Digest*).

- [C12] Zeyu Liu*, **Yizhong Wang***, Jungo Kasai, Hannaneh Hajishirzi, and Noah A. Smith. “Probing Across Time: What Does RoBERTa Know and When?” In: *Proceedings of EMNLP Findings*. 2021.
- [C11] Alon Talmor, Ori Yoran, Amnon Catav, Dan Lahav, **Yizhong Wang**, Akari Asai, Gabriel Ilharco, Hannaneh Hajishirzi, and Jonathan Berant. “MultiModalQA: Complex Question Answering Over Text, Tables, and Images”. In: *Proceedings of ICLR*. 2021.
- [C10] Yue Guo, Wei Qiu, **Yizhong Wang**, and Trevor Cohen. “Automated Lay Language Summarization of Biomedical Scientific Reviews”. In: *Proceedings of AAAI*. 2021.
- [C9] Swabha Swayamdipta, Roy Schwartz, Nicholas Lourie, **Yizhong Wang**, Hannaneh Hajishirzi, Noah A. Smith, and Yejin Choi. “Dataset Cartography: Mapping and Diagnosing Datasets with Training Dynamics”. In: *Proceedings of EMNLP*. 2020.
- [C8] Qianying Liu, Sicong Jiang, Yizhong Wang, and Sujian Li. “LiveQA: A Question Answering Dataset Over Sports Live”. In: *Proceedings of Chinese Computational Linguistics Conference*. 2020.
- [C7] Eric Wallace*, **Yizhong Wang***, Sujian Li, Sameer Singh, and Matt Gardner. “Do NLP Models Know Numbers? Probing Numeracy in Embeddings”. In: *Proceedings of EMNLP-IJCNLP*. 2019.
- [C6] Dheeru Dua, **Yizhong Wang**, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. “DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs”. In: *Proceedings of NAACL-HLT*. 2019. ([NAACL 2019 Most Influential Paper #8 by Paper Digest](#)).
- [C5] **Yizhong Wang**, Sujian Li, and Jingfeng Yang. “Toward Fast and Accurate Neural Discourse Segmentation”. In: *Proceedings of EMNLP*. 2018.
- [C4] **Yizhong Wang**, Kai Liu, Jing Liu, Wei He, Yajuan Lyu, Hua Wu, Sujian Li, and Haifeng Wang. “Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification”. In: *Proceedings of ACL*. 2018.
- [C3] Shuming Ma, Xu Sun, **Yizhong Wang**, and Junyang Lin. “Bag-of-Words as Target for Neural Machine Translation”. In: *Proceedings of ACL*. 2018.
- [C2] **Yizhong Wang**, Sujian Li, and Houfeng Wang. “A Two-Stage Parsing Method for Text-Level Discourse Analysis”. In: *Proceedings of ACL*. 2017. ([Outstanding Paper Award](#)).
- [C1] **Yizhong Wang**, Sujian Li, Jingfeng Yang, Xu Sun, and Houfeng Wang. “Tag-Enhanced Tree-Structured Neural Networks for Implicit Discourse Relation Classification”. In: *Proceedings of IJCNLP*. 2017.

Peer-Reviewed Workshop Papers

- [W2] Joel Jang, Seungone Kim, Bill Yuchen Lin, **Yizhong Wang**, Jack Hessel, Luke Zettlemoyer, Hannaneh Hajishirzi, Yejin Choi, and Prithviraj Ammanabrolu. “Personalized Soups: Personalized Large Language Model Alignment via Post-Hoc Parameter Merging”. In: *Proceedings of NeurIPS 2024 AFM Workshop*. 2024.
- [W1] Wei He, Kai Liu, Jing Liu, Yajuan Lyu, Shiqi Zhao, Xinyan Xiao, Yuan Liu, **Yizhong Wang**, Hua Wu, Qiaoqiao She, et al. “DuReader: A Chinese Machine Reading Comprehension Dataset from Real-World Applications”. In: *Proceedings of ACL MRQA Workshop*. 2018.

Preprints

- [P3] Xinxi Lyu, **Yizhong Wang**, Hannaneh Hajishirzi, and Pradeep Dasigi. “Href: Human response-guided evaluation of instruction following in language models”. In: *arXiv* (2024).
- [P2] Hamish Ivison*, **Yizhong Wang***, Valentina Pyatkin, Nathan Lambert, Matthew Peters, Pradeep Dasigi, Joel Jang, David Wadden, Noah A. Smith, Iz Beltagy, et al. “Camels in a Changing Climate: Enhancing LM Adaptation with Tulu 2”. In: *arXiv* (2023).
- [P1] Aleksandra Piktus, Fabio Petroni, **Yizhong Wang**, Vladimir Karpukhin, Dmytro Okhonko, Samuel Broscheit, Gautier Izacard, Patrick Lewis, Barlas Oğuz, Edouard Grave, Wen-tau Yih, and Sebastian Riedel. “The Web Is Your Oyster: Knowledge-Intensive NLP Against a Very Large Web Corpus”. In: *arXiv* (2022).

Invited Talks

02-03/2025	GaTech / UNC / UMass / UT Austin / UW Madison / Harvard / HKU / HKUST Title: Building a Sustainable Data Foundation for AI
04/10/2024	Princeton University, AI Alignment Group Title: Temporal Alignment of Language Models
02/27/2024	University of Washington, RAIVN Lab Title: Tulu 1, 2, and Maybe 3? Exploring Post-Training Improvement of Open LLMs
06/30/2023	Tsinghua University, NLP Group Title: How Far Can Camels Go? Exploring the State of Instruction Tuning on Open Resources
04/12/2023	Microsoft Research, Knowledge and Language Team Title: Synthetic Instructions for Aligning Language Models
04/01/2023	Baidu Research Title: Instruction Tuning of Language Models: Preliminary Findings
12/29/2022	Alibaba Damo Academy Title: Instruction Tuning of Language Models
11/30/2022	NVIDIA Research Title: Super-NaturalInstructions: Generalization via Declarative Instructions on 1,600+ NLP Tasks
05/29/2019	Peking University, AI Seminar Title: Complex Machine Reading Comprehension with Discrete Reasoning Processes

Teaching

Conference Tutorial

ACL 2025 [Synthetic Data in the Era of Large Language Models](#)
with Vijay Viswanathan, Xiang Yue, Alisa Liu and Graham Neubig

Teaching Assistant

Autumn 2023	CSE 473: Introduction to Artificial Intelligence, University of Washington Head TA Instructor: Hannaneh Hajishirzi
Spring 2022	CSE 481N: Natural Language Processing Capstone, University of Washington Instructor: Noah Smith
Summer 2018	Introduction to Artificial Intelligence, Peking University Instructor: Vincent Ng
Fall 2016 / Fall 2017	Discrete Math, Peking University Instructor: Sujian Li

Guest Lectures

Fall 2024	CS4501: Natural Language Processing, University of Virginia Topic: State of LLM Post-Training Host: Yu Meng
Fall 2024	Natural Language Processing with Deep Learning, Peking University Topic: State of LLM Post-Training Host: Sujian Li
Spring 2024	COMP3361: Natural Language Processing, Hong Kong University Topic: Training Language Models to Follow Instructions Host: Tao Yu
Spring 2023	CS 471/671 NLP: Self-supervised Models, Johns Hopkins University Topic: Instruction Tuning of Large Language Models Host: Daniel Khoshnab
Spring 2022	CSE 481N: Natural Language Processing Capstone, University of Washington Topic: Cross-Task Generalization via Natural Language Host: Noah Smith

Mentoring

Predoctoral Students

01/2024–11/2024	Lj Miranda, PYI@AI2@AI2 → PhD student@Cambridge
07/2023–06/2024	Bowen Zhao, MS student@UW → Startup
01/2024–06/2024	Mickel Liu, Visiting student@UW → PhD student@UW
05/2023–03/2024	Zander Brumbaugh, BS/MS student@UW → Founder of Gamebeast
01/2023–06/2024	Abhika Mishra, BS/MS student@UW → Meta
04/2022–09/2023	Hamish Ivison, PYI@AI2 → PhD student@UW
06/2022–09/2023	Yeganeh Kordi, BS@Tehran Polytechnic → PhD student@Brown

12/2019–04/2021 Zeyu Liu, BS/MS@UW→AI Resident@Meta→PhD student@UT Austin

PhD Students (close collaboration in their early stage)

09/2024–03/2025 Zhiyuan Zeng, UW CSE
06/2022–11/2024 Alisa Liu, UW CSE
09/2022–09/2023 Yushi Hu, UW CSE
09/2019–02/2021 Yue Guo, UW Medical School

Professional Services

Workshop Organization

NeurIPS 2023 Co-organizer of the [Workshop on Instruction Tuning and Instruction Following](#)
ACL 2020 Co-chair of the [ACL 2020 Student Research Workshop](#)

Area Chair

ACL Rolling Review 2024-2025

Reviewer / Program Committee

ACL 2019-25	EMNLP 2019-25	NAACL 2020-25	CoLM 2024
NeurIPS 2021-25	ICLR 2022-24	ICML 2023-24	AAAI 2020-21
AKBC 2020-22	COLING 2022	AACL 2020	

Departmental Service

2023 Fall Co-organizer of UWNLP Seminar Series
2021, 2022, 2023 UW CSE PhD admission student committee member
2022 Fall Co-organizer of UWNLP Retreat