

Junlin Wang

 Email |  junlinwang.com |  IsThatYou |  Google Scholar |  LinkedIn

Education

Duke University

Ph.D. in Computer Science, advised by Prof. Bhuwan Dhingra

Durham, NC

2022 – Present

- Research: long-horizon tasks, agents, continual learning, and test-time scaling.

University of California, Irvine

M.S. in Computer Science; B.S. in Computer Science & Mathematics

Irvine, CA

2016 – 2022

- Research assistant, advised by Prof. Sameer Singh (2018 – 2022).

Industry Experience

Modal

Research Intern

San Francisco, CA

May 2026 – Aug 2026

- Post-training autoresearch agent.
- Platform for agent post-training agent on specialized domains.
- Worked on autoinference. Automatically hill climb inference serving for llms like kimi k3.

Together AI

AI/ML Research Intern

San Francisco, CA

Jan 2026 – Apr 2026

- Built an agentic benchmark-audit system (autobenchaudit.com) that automatically inspects evaluation datasets, surfacing over 35k problems across 168 widely used benchmarks.

Google DeepMind

Student Researcher

New York, NY

May 2025 – Aug 2025

- Research on textual diffusion models.

Together AI

AI/ML Research Intern (part-time)

Remote

Sep 2024 – Apr 2025

- Investigated the efficiency of verifier-free test-time-scaling methods for reasoning models.

Together AI

AI/ML Research Intern

San Francisco, CA

May 2024 – Aug 2024

- Designed multi-agent inference harnesses that combine open-source models to surpass frontier-model quality, publishing two papers on Mixture-of-Agents (ICLR 2025, ICML 2025).

Amazon AWS

Applied Scientist Intern

Santa Clara, CA

May 2023 – Aug 2023

- One of the first to propose budget-aware evaluation of LLM reasoning strategies, resulting in a paper on test-time scaling (EMNLP 2024).

Publications (* equal contribution)

[1] PRO-LONG: Programmatic Memory Enables Long-Horizon Reasoning

Alexis Fox, Junlin Wang, Paul Rosu, Bhuwan Dhingra

arXiv preprint, 2026 [\[paper\]](#)

[2] Vision2Code: A Multi-Domain Benchmark for Evaluating Image-to-Code Generation

Ajay Vikram Periasami, Junlin Wang, Bhuwan Dhingra

In submission, 2026 [\[paper\]](#)

[3] DSGym: A Standardized Framework for Evaluating and Training Data Science Agents

Junlin Wang et al.

ICML 2026 [\[paper\]](#)

[4] How Much Backtracking is Enough? Exploring the Interplay of SFT and RL in Enhancing LLM Reasoning

Hongyi James Cai*, Junlin Wang*, Xiaoyin Chen, Bhuwan Dhingra

COLM 2026 [\[paper\]](#)

- [5] **Improving Model Alignment Through Collective Intelligence of Open-Source Models**
Junlin Wang, Roy Xie, Shang Zhu, Jue Wang, Ben Athiwaratkun, Bhuwan Dhingra, Shuaiwen Leon Song, Ce Zhang, James Zou
ICML 2025 [\[paper\]](#)
- [6] **Think Deep, Think Fast: Investigating Efficiency of Verifier-Free Inference-Time-Scaling Methods**
Junlin Wang, Shang Zhu, Jon Saad-Falcon, Ben Athiwaratkun, Qingyang Wu, Jue Wang, Shuaiwen Leon Song, Ce Zhang, Bhuwan Dhingra, James Zou
ICML AdaptFM Workshop 2026 [\[paper\]](#)
- [7] **Knowing When to Stop: Dynamic Context Cutoff for Large Language Models**
Roy Xie, Junlin Wang, Paul Rosu, Chunyuan Deng, Bolun Sun, Zihao Lin, Bhuwan Dhingra
NeurIPS 2025 [\[paper\]](#)
- [8] **Staircase Streaming for Low-Latency Multi-Agent Inference**
Junlin Wang, Jue Wang, Zhen Xu, Ben Athiwaratkun, Bhuwan Dhingra, Ce Zhang, James Zou
ICML AdaptFM Workshop 2026 [\[paper\]](#)
- [9] **Mixture-of-Agents Enhances Large Language Model Capabilities**
Junlin Wang, Jue Wang, Ben Athiwaratkun, Ce Zhang, James Zou
ICLR 2025 [\[paper\]](#)
- [10] **Token Economies: Budget-Aware Evaluation of LLM Reasoning Strategies**
Junlin Wang, Siddharth Jain, Dejiao Zhang, Baishakhi Ray, Varun Kumar, Ben Athiwaratkun
EMNLP 2024 [\[paper\]](#)
- [11] **ReCaLL: Membership Inference via Relative Conditional Log-Likelihoods**
Roy Xie, Junlin Wang, Ruomin Huang, Minxing Zhang, Rong Ge, Jian Pei, Neil Zhenqiang Gong, Bhuwan Dhingra
EMNLP 2024 [\[paper\]](#)
- [12] **Adversarial Math Word Problem Generation**
Roy Xie, Chengxuan Huang, Junlin Wang, Bhuwan Dhingra
EMNLP Findings 2024 [\[paper\]](#)
- [13] **Raccoon: Prompt Extraction Benchmark of LLM-Integrated Applications**
Junlin Wang*, Tianyi Yang*, Roy Xie, Bhuwan Dhingra
ACL Findings 2024 [\[paper\]](#)
- [14] **NeuroComparatives: Neuro-Symbolic Distillation of Comparative Knowledge**
Phillip Howard*, Junlin Wang*, Vasudev Lal, Gadi Singer, Yejin Choi, Swabha Swayamdipta
NAACL Findings 2023 [\[paper\]](#)
- [15] **GAP-Gen: Guided Automatic Python Code Generation**
Junlin Wang*, Yurun Song*, Junchen Zhao*, Ian Harris
EACL Workshop 2022 [\[paper\]](#)
- [16] **Maestro: A Gamified Platform for Teaching AI Robustness**
Margarita Geleta, Jiachen Xu, Manikanta Loya, Junlin Wang, Sameer Singh, Zhou Li, Sergio Gago-Masague
EAAI 2022 [\[paper\]](#)
- [17] **Gradient-based Analysis of NLP Models is Manipulable**
Junlin Wang*, Jens Tuyls*, Eric Wallace, Sameer Singh
EMNLP Findings 2020 [\[paper\]](#)
- [18] **AllenNLP Interpret: A Framework for Explaining Predictions of NLP Models**
Eric Wallace, Jens Tuyls, Junlin Wang, Sanjay Subramanian, Matt Gardner, Sameer Singh
EMNLP 2019 [\(Best Demo Award\)](#) [\[paper\]](#) [\[landing page\]](#) [\[demo\]](#)

Honors and Awards

2026	Best public score, ARC-AGI-3
2019	Best Demo Award, EMNLP 2019
2019	Best Visualization, DataFest 2019
2018	E-SURP Fellowship, Edwards Lifesciences Center, UCI
2018–2020	ICS Honors Program , UCI
2017	3rd Place, Microsoft Coding Competition at UCI
2017	Best Project and Development Practices, BeachHacks
2017	1st Place, 1st Tippers IoT Hackathon

Invited Talks

Mixture-of-Agents Enhances Large Language Model Capabilities *Oct 2024*
Microsoft.

Feedback Mini-Batching: Automatically Detecting Data Artifacts and Robust Training *Sep 2019*
Lab Week Talk, Comcast Applied AI Research Lab, Washington, DC.

Academic Service

Reviewer **ICML**, 2025-2026
 NeurIPS, 2024–2026
 ICLR, 2024–2026
 ACL ARR, Dec 2023, Feb 2024, Jun 2024, Oct 2024