

Yangjun Ruan

UNIVERSITY OF TORONTO & VECTOR INSTITUTE

(+1) 650-441-9054
www.cs.toronto.edu/~yjruan
yjruan@cs.toronto.edu

EDUCATION

Department of Computer Science, **University of Toronto**
Ph.D. Student

Sep. 2020 - Present

- Affiliated with Vector Institute & Machine learning group
- Advisors: Chris J. Maddison & Jimmy Ba

Department of Information Science & Electronic Engineering, **Zhejiang University**
B.Eng., Information Engineering

Sep. 2016 - Jun. 2020

- GPA: 94.1/100, Major: 94.8/100, **Rank: 1/140** (three consecutive years)
- Graduated with the highest honor (CHU Kochen Scholarship)

RESEARCH VISITS

Department of Computer Science, **Stanford University**
Visiting Student Researcher

Nov. 2023 - Present

- Advisor: Tatsunori Hashimoto

Department of Computer Science, **University of California, Los Angeles**
Visiting Research Intern

Jul. 2019 - Sep. 2019

- Cross-disciplinary Scholars in Science and Technology (**CSST**)
- Advisor: Cho-Jui Hsieh

RESEARCH INTERESTS

My research focuses on scaling, evaluating, and aligning language models and agents, especially as they approach or exceed super-human performance levels.

PUBLICATIONS

Preprints

- [Observational Scaling Laws and the Predictability of Language Model Performance](#)
Yangjun Ruan, Chris J. Maddison, Tatsunori Hashimoto.
arXiv Preprint 2405.10938

Conference papers

- [Identifying the Risks of LM Agents with an LM-Emulated Sandbox](#)
Yangjun Ruan*, Honghua Dong*, Andrew Wang, Silviu Pitis, Yongchao Zhou, Jimmy Ba, Yann Dubois, Chris J. Maddison, Tatsunori Hashimoto.
International Conference on Learning Representations (ICLR), 2023. [**Spotlight**]
- [Weighted Ensemble Self-Supervised Learning](#)
Yangjun Ruan, Saurabh Singh, Warren R. Morningstar, Alexander A. Alemi, Sergey Ioffe, Ian Fischer, Joshua V. Dillon.
International Conference on Learning Representations (ICLR), 2023.
- [Augment with Care: Contrastive Learning for Combinatorial Problems](#)
Haonan Duan, Pashootan Vaezipoor, Max B. Paulus, **Yangjun Ruan**, Chris J. Maddison.
International Conference on Machine Learning (ICML), 2022.
- [Optimal Representations for Covariate Shift](#)
Yangjun Ruan*, Yann Dubois*, Chris J. Maddison.
International Conference on Learning Representations (ICLR), 2022.
- [Improving Lossless Compression Rates via Monte Carlo Bits-Back Coding](#)
Yangjun Ruan*, Karen Ullrich*, Daniel Severo*, James Townsend, Ashish Khisti, Arnaud Doucet, Alireza Makhzani, Chris J. Maddison.
International Conference on Machine Learning (ICML), 2021. [**Long talk**]

- [Learning to Learn by Zeroth-Order Oracle](#)
Yangjun Ruan, Yuanhao Xiong, Sashank Reddi, Sanjiv Kumar, Cho-Jui Hsieh.
International Conference on Learning Representations (ICLR), 2020.
- [FastSpeech: Fast, Robust and Controllable Text to Speech](#)
Yi Ren*, Yangjun Ruan*, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, Tie-Yan Liu.
Advances in Neural Information Processing Systems (NeurIPS), 2019.
- [Data Transmission in Mobile Edge Networks: Whether and Where to Compress?](#)
Jinke Ren*, Yangjun Ruan*, Guanding Yu.
IEEE Communications Letters 23 (3), 490-493.

Workshop papers

- [Calibrating Language Models via Augmented Prompt Ensembles](#)
Mingjian Jiang*, Yangjun Ruan*, Sicong Huang, Saifei Liao, Silviu Pitis, Roger Grosse, Jimmy Ba
ICML Workshop on Deployment Challenges for Generative AI, 2023.

Note: * above denotes equal contribution.

RESEARCH EXPERIENCE

Stanford University , Visiting Student Researcher Advisor: Tatsunori Hashimoto Topic: language models, agent, evaluation, scalable oversight	Palo Alto Nov. 2023 - Present
University of Toronto & Vector Institute , Research Assistant Advisor: Chris J. Maddison, Jimmy Ba Topic: language models, agent, evaluation	Toronto Oct. 2022 - Present
Google Research , Student Researcher Advisor: Ian Fischer, Joshua V. Dillon Topic: self-supervised learning, ensemble method	Mountain View Jun. 2022 - Sep. 2022
University of Toronto & Vector Institute , Research Assistant Advisor: Chris J. Maddison Topic: representation learning, distribution shift, neural compression	Toronto Jul. 2020 - Mar. 2022
Microsoft Research Asia , Research Intern Advisor: Li Dong, Furu Wei Topic: implicit deep learning methods, Transformer model	Beijing Nov. 2019 - Jun. 2020
University of California Los Angeles , Visiting Research Intern Advisor: Cho-Jui Hsieh Topic: learning to learn, zeroth-order optimization, adversarial robustness	Los Angeles Jul. 2019 - Sep. 2019
Zhejiang University , Research Assistant Advisor: Zhou Zhao, Tao Qin Topic: non-autoregressive seq-to-seq model	Hangzhou Feb. 2019 - Jun. 2019

TALKS

ToolEmu: Identifying the Risks of LM Agents with an LM-Emulated Sandbox	
• AI TIME Special Talk Forum	Jan. 2024
• Vector Institute, AI Safety Seminar	Dec. 2023
• Google Research, Robustness Talk Series	Nov. 2023
• Toronto Data Workshop	Oct. 2023
Optimal Representations for Covariate Shift	
• Google Research	Aug. 2022
• CMU, OOD Robustness and Generalization Seminar	Jun. 2022
Monte Carlo Bits-Back Coding	
• ICML [Long talk]	Jun. 2021

- ICLR Neural Compression Workshop [**Oral**] May. 2021

SERVICES

I served as

- Conference reviewer: NeurIPS (20'-), ICLR (21'-), ICML (21'-)
- Workshop reviewer: NeurIPS DGMs Applications Workshop (21'), NeurIPS Pretraining Workshop (22'), ICLR Mathematical and Empirical Understanding of Foundation Models Workshop (23')

AWARDS & HONORS

- Ontario Graduate Scholarship Jul. 2023
 - Outstanding Reviewer for ICML 2022 Jul. 2022
 - DiDi Graduate Student Award Dec. 2021
 - Computer Science 50th Anniversary Graduate Scholarship Dec. 2020
 - CHU Kochen Scholarship Oct. 2019
 - **Highest** scholarship for only **top 12** undergraduates at Zhejiang University
 - National Scholarship (top 1.5%) Oct. 2017, 2018, 2019
 - Cross-disciplinary Scholars in Science and Technology (CSST), UCLA Jul. 2019
 - CSST Best Research Presenter, UCLA Sep. 2019
 - Meritorious Winner, Interdisciplinary Contest in Modeling (ICM) May. 2018
-