

# Yiren Jian

## Member of Technical Staff

OpenAI

✉ yiren.jian.gr@dartmouth.edu | 🏠 yiren-jian.github.io | 📁 yiren-jian

## Professional Experience

### Member of Technical Staff, OpenAI

Build safe and beneficial AGI—AI systems that are generally smarter than humans.

San Francisco, CA

Mar. 2025 - Now

### Research Scientist, TikTok

Develop multimodal language models for math and STEM domains.

Bellevue, WA

Sept. 2024 - Mar. 2025

Manager: Zhenheng Yang

### Research Scientist, ByteDance

Develop open-sourced visual language models for multimodal understanding.

Bellevue, WA

Feb. 2024 - Sept. 2024

Manager: Hongxia Yang

### Research Intern, ByteDance

Develop efficient generative visual-language models based on LLM.

Bellevue, WA

June. 2023 - Sept. 2023

Advisor: Yunzhe Tao

### Research Intern, Snap Research

Develop neural speakers for affective image captioning.

Palo Alto, CA

June. 2022 - Sept. 2022

Advisor: Panos Achlioptas

### Research Intern, NEC Labs America

Develop physical-augmented ML models for precision immunotherapy.

Princeton, NJ (remote)

June. 2021 - Sept. 2021

Advisor: Martin Renqiang Min

## Education

### Dartmouth College

Doctor of Philosophy - Computer Science

Hanover, NH, USA

Sept. 2018 - Jan. 2024

Advisor: Soroush Vosoughi

### The George Washington University

Master of Science - Biophysics

Washington DC, USA

Sept. 2015 - May. 2017

Advisor: Chen Zeng

### Huazhong University of Science and Technology

Bachelor of Science - Physics

Wuhan, China

Sept. 2011 - May. 2015

## Publications

### Pretrained Image-Text Models are Secretly Video Captioners

[NAACL 2025, 🌟 [oral presentation](#)] The Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics

Chunhui Zhang, **Yiren Jian**, Zhongyu Ouyang, Soroush Vosoughi

### Working Memory Identifies Reasoning Limits in Language Models

[EMNLP 2024] The Conference on Empirical Methods in Natural Language Processing

Chunhui Zhang, **Yiren Jian**, Zhongyu Ouyang, Soroush Vosoughi

### InfiMM-WebMath-40B: Advancing Multimodal Pre-Training for Enhanced Mathematical Reasoning

preprint

Xiaotian Han<sup>#</sup> and **Yiren Jian**<sup>#</sup> and Xuefeng Hu<sup>#</sup> and Haogeng Liu<sup>#</sup> and Yiqi<sup>#</sup> Wang and Qihang Fan and Yang Ai and Huaibo Huang and Ran He and Zhenheng Yang and Quanzeng You

## **InfiMM: Advancing Multimodal Understanding with an Open-Sourced Visual Language Model**

[Findings of ACL 2024] Findings of the Association for Computational Linguistics

Haogeng Liu, Quanzeng You, Yiqi Wang, Xiaotian Han, Bohan Zhai, Yongfei Liu, Wentao Chen, **Yiren Jian**, Yunzhe Tao, Jianbo Yuan, Ran He, Hongxia Yang

## **Expedited Training of Visual Conditioned Language Generation via Redundancy Reduction**

[ACL 2024, ✨ *oral presentation*] The 62nd Annual Meeting of the Association for Computational Linguistics

**Yiren Jian**, Tingkai Liu, Yunzhe Tao, Chunhui Zhang, Soroush Vosoughi, Hongxia Yang

## **GEM: Generating Engaging Multimodal Content**

[IJCAI 2024] International Joint Conference on Artificial Intelligence

Chongyang Gao, **Yiren Jian**, Natalia Denisenko, Soroush Vosoughi, V.S. Subrahmanian

## **Efficient and Effective Learning of Foundational Large Multi-Modal Models**

Dartmouth College PhD Dissertations, 2024

**Yiren Jian**

## **RNet: a network strategy to predict RNA binding preferences**

Briefings in Bioinformatics, 2024

Haoquan Liu, **Yiren Jian**, Jinxuan Hou, Ceng Zeng, Yunjie Zhao

## **Knowledge from Large-Scale Protein Contact Prediction Models Can Be Transferred to the Data-Scarce RNA Contact Prediction Task**

[ICPR 20224] International Conference on Pattern Recognition, 2024

**Yiren Jian**<sup>†</sup>, Chongyang Gao, Chen Zeng, Yunjie Zhao, Soroush Vosoughi<sup>†</sup>

## **Bootstrapping Vision-Language Learning with Decoupled Language Pre-training**

[NeurIPS 2023, ✨ *spotlight*] Advances in Neural Information Processing Systems

**Yiren Jian**, Chongyang Gao, Soroush Vosoughi

## **Evaluation of DNA-protein complex structures using the deep learning method**

Physical Chemistry and Chemical Physics, 2023

Chengwei Zeng<sup>#</sup>, **Yiren Jian**<sup>#</sup>, Chen Zhuo, Anbang Li, Chen Zeng, Yunjie Zhao

## **Evaluating Native-like Structures of RNA-protein Complexes Through the Deep Learning Method**

Nature Communications, 2023

Chengwei Zeng<sup>#</sup>, **Yiren Jian**<sup>#</sup>, Soroush Vosoughi, Chen Zeng, Yunjie Zhao

## **Non-Linguistic Supervision for Contrastive Learning of Sentence Embeddings**

[NeurIPS 2022] Advances in Neural Information Processing Systems

**Yiren Jian**, Chongyang Gao, Soroush Vosoughi

## **T-Cell Receptor-Peptide Interaction Prediction with Physical Model Augmented Pseudo-Labeling**

[KDD 2022, ✨ *oral presentation*] In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

**Yiren Jian**, Erik Kruus, Martin Renqiang Min

## **Embedding Hallucination for Few-shot Language Fine-tuning**

[NAACL 2022] In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

**Yiren Jian**, Chongyang Gao, Soroush Vosoughi

## **Contrastive Learning for Prompt-based Few-shot Language Learners**

[NAACL 2022] In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

**Yiren Jian**, Chongyang Gao, Soroush Vosoughi

## **Label Hallucination for Few-Shot Classification**

[AAAI 2022] Proceedings of the 36th AAAI Conference on Artificial Intelligence

**Yiren Jian**, Lorenzo Torresani

## **MetaPix: Domain Transfer for Semantic Segmentation by Meta Pixel Weighting**

Image and Vision Computing, 2021

**Yiren Jian**, Chongyang Gao

## **Task Meta-Transfer from Limited Parallel Labels**

Meta-Learning Workshop at Neural Information Processing Systems, 2020

**Yiren Jian**, Karim Ahmed, Lorenzo Torresani

## **DIRECT: RNA Contact Predictions by Integrating Structural Patter**

BMC Bioinformatics, 2019

**Yiren Jian**, Xiaonan Wang, Jiadi Qiu, Huiwen Wang, Zhichao Liu, Yunjie Zhao, Chen Zeng

## **Trace, Machine Learning of Signal Images for Trace-Sensitive Mass Spectrometry: A Case Study from Single-Cell Metabolomics**

Analytical Chemistry, 2019

Zhichao Liu, Erika P. Portero, **Yiren Jian**, Yunjie Zhao, Rosemary M. Onjiko, Chen Zeng, Peter Nemes

## **Design of Tat-activated Cdk9 Inhibitor**

Journal of Peptide Research and Therapeutics, 2019

Yunjie Zhao, Hao Chen, Chenghang Du, **Yiren Jian**, Haotian Li, Yi Xiao, Mohammed Saifuddin, Fatah Kashanchi, and Chen Zeng

## **Rbind: Computational Network Method to Predict the Binding Sites of RNA Molecules**

Bioinformatics, 2018

Kaili Wang<sup>#</sup>, **Yiren Jian**<sup>#</sup>, Huiwen Wang, Chen Zeng and Yunjie Zhao

## **Computational Study of Non-catalytic T-loop Pocket on CDK Proteins for Drug Development**

Chinese Physics B, 2017

Huiwen Wang, Kaili Wang, Zeyu Guan, **Yiren Jian**, Ya Jia, Fatah Kashanchi, Chen Zeng, Yunjie Zhao

## **Network Analysis Reveals the Recognition Mechanism for Dimer Formation of Bulb-type Lectins**

Scientific Report, 2017

Yunjie Zhao<sup>#</sup>, **Yiren Jian**<sup>#</sup>, Zhichao Liu, Hang Liu, Qin Liu, Chanyou Chen, Zhangyong Li, Lu Wang, H. Howie Huang, Chen Zeng

## **Patents**

---

**T-cell receptor repertoire selection prediction with physical model augmented pseudo-labeling** (w/ NEC Labs America)

## **Teaching Experience**

---

2019, 2021	<b>Graduate Teaching Assistant</b> , Deep Learning (graduate-level course)	Dartmouth
2021-2023	<b>Graduate Teaching Assistant</b> , Machine Learning (graduate-level course)	Dartmouth
2018	<b>Graduate Teaching Assistant</b> , Machine Learning (graduate-level course)	Dartmouth
2015-2016	<b>Graduate Teaching Assistant</b> , University Physics (undergraduate course)	GWU