

Dongfu Jiang

Updated June 20, 2025

✉ Email: dongfu.jiang@uwaterloo.ca

🏠 Home Page: jdf-prog.github.io

🐙 GitHub: [jdf-prog](https://github.com/jdf-prog)

🌐 LinkedIn: [dongfu-jiang](https://www.linkedin.com/in/dongfu-jiang)

Research interests

Natural Language Processing; Large Language Models; Evaluation; Multimodality Reasoning;

Education

University of Waterloo

PhD in Computer Science

Advisor: Prof. Wenhua Chen.

Waterloo, Canada

Sep 2023 – Present

Zhejiang University

BE in Computer Science

Advisor: Prof. Zhou Zhao. GPA: 3.97/4.00.

Hangzhou, China

Sep 2019 – June 2023

Open Source Projects

VerlTool: A unified and easy-to-extend tool-agent training framework based on verl.

Project Leader: [Dongfu Jiang](#)

Core Contributors: Zhuofeng Li, Yi Lu, Zhiheng Lv, Ping Nie

Advisors: Chao Du, Tianyu Pang, Wenhua Chen

🐙 [TIGER-AI-Lab/verl-tool](https://github.com/TIGER-AI-Lab/verl-tool) ★ 250

LLM-Engines: A Unified and Parallel Inference Engine for Large Language Models

[Dongfu Jiang](#)

🐙 [jdf-prog/LLM-Engines](https://github.com/jdf-prog/LLM-Engines) ★ 48

Publications

StructEval: Benchmarking LLMs' Capabilities to Generate Structural Outputs

Jialin Yang*, [Dongfu Jiang*](#), Lipeng He, Sherman Siu, Yuxuan Zhang, Disen Liao, Zhuofeng Li, Huaye Zeng, Yiming Jia, Haozhe Wang, Benjamin Schneider, Chi Ruan, Wentao Ma, Zhiheng Lyu, Yifei Wang, Yi Lu, Quy Duc Do, Ziyang Jiang, Ping Nie, Wenhua Chen

Under review of EMNLP 2025.

🐙 [TIGER-AI-Lab/StructEval](https://github.com/TIGER-AI-Lab/StructEval) ★ 5

QuickVideo: Real-Time Long Video Understanding with System Algorithm Co-Design

Benjamin Schneider*, [Dongfu Jiang*](#), Chao Du, Tianyu Pang, Wenhua Chen

Under review of NeurIPS 2025.

🐙 [TIGER-AI-Lab/QuickVideo](https://github.com/TIGER-AI-Lab/QuickVideo) ★ 55

General-Reasoner: Advancing LLM Reasoning Across All Domains

Xueguang Ma*, Qian Liu*, Dongfu Jiang, Ge Zhang, Zejun Ma, Wenhui Chen

Under review of NeurIPS 2025.

 [TIGER-AI-Lab/General-Reasoner](#) ★ 141

ACECODER: Acing Coder RL via Automated Test-Case Synthesis

Huaye Zeng*, Dongfu Jiang*, Haozhe Wang, Ping Nie, Xiaotong Chen, and Wenhui Chen

ACL, 2025.

 [TIGER-AI-Lab/AceCoder](#) ★ 85

MEGA-Bench: Scaling Multimodal Evaluation to over 500 Real-World Tasks

Jiacheng Chene*, Tianhao Liang*, Sherman Siue*, Zhengqing Wang, Kai Wang, Yubo Wang, Yuansheng Ni, Wang Zhu, Ziyang Jiang, Bohan Lyu, Dongfu Jiang, Xuan He, Yuan Liu, Hexiang Hu, Xiang Yue, Wenhui Chen

ICLR, 2025.

 [TIGER-AI-Lab/MEGA-Bench](#) ★ 68

VideoScore: Building Automatic Metrics to Simulate Fine-grained Human Feedback for Video Generation

Xuan He*, Dongfu Jiang*, Ge Zhang, Max Ku, Achint Soni, Sherman Siu, Hao-nan Chen, Abhramil Chandra, and Ziyang Jiang, Aaran Arulraj, Kai Wang, Quy Duc Do, Yuansheng Ni, Bohan Lyu, Yaswanth Narsupalli, Rongqi Fan, Zhiheng Lyu, Yuchen Lin, Wenhui Chen

EMNLP, 2024.

 [TIGER-AI-Lab/VideoScore](#) ★ 91

WildVision: Evaluating Vision-Language Models in the Wild with Human Preferences

Yujie Lu, Dongfu Jiang, Wenhui Chen, William Yang Wang, Yejin Choi, and Bill Yuchen Lin

NeurIPS 2024 Dataset and Benchmark Track, 2024.

GenAI Arena: An Open Evaluation Platform for Generative Models

Dongfu Jiang*, Max W.F. Ku*, Tianle Li*, Yuansheng Ni, Shizhuo Sun, Rongqi Fan, and Wenhui Chen

MANTIS: Interleaved Multi-Image Instruction Tuning

Dongfu Jiang, Xuan He, Huaye Zeng, and Cong Wei, Max Ku, Qian Liu, Wenhui Chen

Transactions on Machine Learning Research, 2024.

 [TIGER-AI-Lab/Mantis](#) ★ 218

VIEScore: Towards Explainable Metrics for Conditional Image Synthesis Evaluation

Max Ku, Dongfu Jiang, Cong Wei, Xiang Yue, Wenhui Chen

ACL, 2024.

 [TIGER-AI-Lab/VieScore](#) ★ 45

MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI

Xiang Yue, Yuansheng Ni, Kai Zhang, and Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, Wenhui Chen

CVPR *Oral*, 2024.

 [MMMU-Benchmark/MMMU](#) ★ 447

TIGERScore: Towards Building Explainable Metric for All Text Generation Tasks

Dongfu Jiang^{*}, Yishan Li^{*}, Ge Zhang, Wenhao Huang, Bill Yuchen Lin, Wenhui Chen

Transactions on Machine Learning Research, 2024.

 [TIGER-AI-Lab/TIGERScore](#) ★ 31

LLM-Blender: Ensembling Large Language Models with Pairwise Ranking and Generative Fusion

Dongfu Jiang, Xiang Ren, Bill Yuchen Lin

ACL, 2023.

 [yuchenlin/LLM-Blender](#) ★ 947

Experience

Research Intern at Allen Institute of AI

Seattle, US

Advisors: (Bill) Yuchen Lin, Qianhui Qin

June 2024 – Sep 2024

Active Learning with Verbalized Human Feedback

Research associate at SeaAI

Singapore (Remote)

Advisors: Qian Liu, Tianyu Pang, Chao Du

Feb 2024 – Today

MANTIS: Interleaved Multi-Image Instruction Tuning

Research Intern at University of Southern California

US (Remote)

Advisors: Xiang Ren, (Bill) Yuchen Lin

Mar 2022 – Mar 2023

LLM-Blender: Ensembling Large Language Models with Pairwise Ranking and Generative Fusion

Honors and Scholarships

National Scholarship (ZJU, *top 1%*)

2020

Zhejiang University First-Class Scholarship (ZJU, *top 3%*)

2020, 2021

