

Lawrence Jang

CONTACT

EMAIL: ljang@cs.cmu.edu WEBSITE: lawrenceckjang.github.io

RESEARCH INTERESTS

Aligning human preferences and instilling human capabilities within LLM agents; Reproducible evaluation and benchmarking LLM agents; LLM agents for automating everyday human computer tasks

EDUCATION

Carnegie Mellon University, Pittsburgh, PA, United States

M.S. Machine Learning, GPA: 4.00

Jan 2024 - Dec 2024

- Advisor(s): Dr. Ruslan Salakhutdinov, Dr. Katerina Fragkiadaki
- Relevant courses: Advanced Deep Learning, Probabilistic Graphical Models, Theoretical and Empirical Foundations of Modern Machine Learning, Probability and Mathematical Statistics, Advanced Machine Learning Theory

Carnegie Mellon University, Pittsburgh, PA, United States

B.Sc. Statistics and Machine Learning, GPA: 3.73

Aug 2019 - Dec 2023

- Research advisor(s): Dr. Ron Yurko
- Relevant courses: Convex Optimization, Distributed Machine Learning with Large Datasets, Deep Reinforcement Learning, Non-Parametric Learning, Modern Regression, Multilevel Modeling

PREPRINTS

TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks

Frank F. Xu, Yufan Song, Boxuan Li, Yuxuan Tang, Kritanjali Jain, Mengxue Bao, Zora Z. Wang, Xuhui Zhou, Zhitong Guo, Murong Cao, Mingyang Yang, Hao Yang Lu, Amaad Martin, Zhe Su, Leander Maben, Raj Mehta, Wayne Chi, **Lawrence Jang**, Yiqing Xie, Shuyan Zhou, Graham Neubig
ArXiv Preprint

The BrowserGym Ecosystem for Web Agent Research

Thibault Le Sellier de Chezelles, Maxime Gasse, Alexandre Lacoste, Massimo Caccia, Alexandre Drouin, Léo Boisvert, Megh Thakkar, Tom Marty, Rim Assouel, Sahar Omidi Shayegan, **Lawrence Jang**, Xing Han Lù, Ori Yoran, Dehan Kong, Frank F. Xu, Siva Reddy, Graham Neubig, Quentin Cappart, Russ Salakhutdinov, Nicolas Chapados
Under review, TMLR

PUBLICATIONS

VideoWebArena: Evaluating Long Context Multimodal Agents with Video Understanding Web Tasks

Lawrence Jang, Yinheng Li, Charles Ding, Justin Lin, Paul Liang, Dan Zhao, Rogerio Bonatti, Kazuhito Koishida (*work done while at Microsoft*)

Workshop on Open-World Agents at NeurIPS 2024

Under review, ICLR 2025

Windows Agent Arena: Evaluating Multi-Modal OS Agents at Scale

Rogerio Bonatti, Dan Zhao, Francesco Bonacci, Dillon Dupont, Sara Abdali, Yinheng Li, Yadong Lu, Justin Wagle, Kazuhito Koishida, Arthur Bucker, **Lawrence Jang**, Zack Hui (*work done while at Microsoft*)

Workshop on Open-World Agents at NeurIPS 2024

Towards Safe and Trustworthy Agents Workshop at NeurIPS 2024 (Oral)

Under review, ICLR 2025

VLM Agents Generate Their Own Memories: Distilling Experience into Embodied Programs

Gabriel Sarch, **Lawrence Jang**, Michael J. Tarr, William W. Cohen, Kenneth Marino, Katerina Fragkiadaki

NeurIPS 2024 (Spotlight)

MMoE: Enhancing Multimodal Models with Mixtures of Multimodal Interaction Experts

Haofei Yu*, Zhengyang Qi*, **Lawrence Jang***, Ruslan Salakhutdinov, Louis-Philippe Morency, Paul Pu Liang

EMNLP 2024

VisualWebArena: Evaluating Multimodal Agents on Realistic Visual Web Tasks

Jing Yu Koh, Robert Lo*, **Lawrence Jang***, Vikram Duvvur*, Ming Chong Lim*, Po-Yu Huang*, Graham Neubig, Shuyan Zhou, Ruslan Salakhutdinov, Daniel Fried

ACL 2024

LLM Agents Workshop at ICLR 2024

ACADEMIC RESEARCH EXPERIENCE	<p>Machine Learning Department, Carnegie Mellon University <i>Research Assistant</i> August 2023 - present</p> <ul style="list-style-type: none"> • Conducted research under Dr. Ruslan Salakhutdinov on multimodal agentic systems and benchmarks. • Conducted research under Dr. Katerina Fragkiadaki on retrieval augmented multimodal agents with self-generating in-context memories. • Collaborated with ServiceNow Research to work on BrowserGym: A Gym Environment for Web Task Automation. • Conducted research under Dr. Ron Yurko on modeling pre-snap NFL defender blitz probabilities. 										
PROFESSIONAL EXPERIENCE	<p>Microsoft <i>Research Intern</i> May 2024 - Aug 2024</p> <ul style="list-style-type: none"> • Helped develop and publish Windows Agent Arena, a scalable OS platform for testing and benchmarking of multi-modal AI agents. • Created VideoWebArena, a video-based benchmark intended to test performance of long-context multimodal agents. <p>Chicago Cubs <i>Research and Development Intern</i> May 2022 - Aug 2022</p> <ul style="list-style-type: none"> • Modeled multivariate joint batted ball distributions with copula models to automate a synthetic data generation of a player's batted ball profile. <p>Major League Baseball Commissioner's Office <i>Data Science Intern</i> May 2021 - Aug 2021</p> <ul style="list-style-type: none"> • Conducted data analysis on spin-rate trends in investigation of foreign substance usage. • Created new high-sensor camera datasets and blog writeups regarding new MLB data innovations. <p>FormFree <i>Data Analytics Intern</i> May 2020 - Aug 2020</p> <ul style="list-style-type: none"> • Identified and manually labeled raw data logs of financial transactions to improve proprietary NLP model. • Created and visualized company financial datasets. 										
HONORS AND AWARDS	<table border="0" style="width: 100%;"> <tr> <td>Carnegie Mellon University Machine Learning Department TA of the Year</td> <td style="text-align: right;">2024</td> </tr> <tr> <td>Carnegie Mellon University Statistics Department Undergraduate TA of the Year</td> <td style="text-align: right;">2023</td> </tr> <tr> <td>CMU Undergraduate University Honors</td> <td style="text-align: right;">2023</td> </tr> <tr> <td>Dean's List (High Honors)</td> <td style="text-align: right;">2021, 2022, 2023</td> </tr> </table>	Carnegie Mellon University Machine Learning Department TA of the Year	2024	Carnegie Mellon University Statistics Department Undergraduate TA of the Year	2023	CMU Undergraduate University Honors	2023	Dean's List (High Honors)	2021, 2022, 2023		
Carnegie Mellon University Machine Learning Department TA of the Year	2024										
Carnegie Mellon University Statistics Department Undergraduate TA of the Year	2023										
CMU Undergraduate University Honors	2023										
Dean's List (High Honors)	2021, 2022, 2023										
TEACHING EXPERIENCE	<p>Carnegie Mellon University <i>Teaching Assistant</i> Aug 2022 - present</p> <ul style="list-style-type: none"> • 10-703 Deep Reinforcement Learning (Fall '22, Spring '23, Fall '23, Fall '24) • 10-708 Probabilistic Graphical Models (Fall '23, Spring '24) • 11-777 Multimodal Machine Learning (Fall '24) • 36-402 Modern Regression (Spring '23, Spring '24) • 36-401 Modern Regression (Fall '23) 										
VOLUNTEERING AND SERVICE	<table border="0" style="width: 100%;"> <tr> <td>Moderator for London Machine Learning Meetup: LLaVA</td> <td style="text-align: right;">2024</td> </tr> <tr> <td>CMU MSML Admissions Committee</td> <td style="text-align: right;">2023</td> </tr> <tr> <td>CMU Paths to AI Research Mentor</td> <td style="text-align: right;">2024-Present</td> </tr> <tr> <td>Founder of CMU Korean-American Scientists and Engineers Association Chapter</td> <td style="text-align: right;">2022-Present</td> </tr> <tr> <td>CMU Club Basketball Head Coach</td> <td style="text-align: right;">2022-Present</td> </tr> </table>	Moderator for London Machine Learning Meetup: LLaVA	2024	CMU MSML Admissions Committee	2023	CMU Paths to AI Research Mentor	2024-Present	Founder of CMU Korean-American Scientists and Engineers Association Chapter	2022-Present	CMU Club Basketball Head Coach	2022-Present
Moderator for London Machine Learning Meetup: LLaVA	2024										
CMU MSML Admissions Committee	2023										
CMU Paths to AI Research Mentor	2024-Present										
Founder of CMU Korean-American Scientists and Engineers Association Chapter	2022-Present										
CMU Club Basketball Head Coach	2022-Present										
TECHNICAL SKILLS	<p>PROGRAMMING: Python, R, SQL</p> <p>MACHINE LEARNING FRAMEWORKS: PyTorch, Tensorflow, HuggingFace Transformers, scikit-learn, wandb, numpy, pandas, matplotlib, ggplot2, tidyverse</p>										