



Michael Sun

✉ msun415@{stanford, mit}.edu

📞 650-285-9287

Profile

My research interests span machine learning over interconnected structures and systems. I enjoy working with domain experts to enable transformative applications. I am applying AI to designing better graph-structured objects — which represent molecules, social networks, materials, buildings, etc. — by learning to reverse engineer existing ones. I want to build an all-in-one platform with solutions for graph-based design used by experts across many domains that transfer knowledge across real-world problems. Give the AI a few examples of what you want, and it will return you something better. I believe this AI engineer will be the last invention we ever need.

I graduated from Stanford with BS Honors in MCS and MS CS in 06/2022. My concentrations for both were in Artificial Intelligence. Prior to deciding to pursue a PhD, I also developed strong industry expertise in machine learning. In my spare time, I built end-to-end full-stack web/mobile applications, filling in missing knowledge gaps and self-learning all tech stacks used. I love tackling hard problems that require creative solutions. Got an interesting opportunity for me? I'm always open to exciting opportunities in research and applied AI.

Education

PhD Computer Science, Massachusetts Institute of Technology

September 2023 — Present

I am an EECS PhD student at MIT and a Graduate RA within MIT Computer Science & Artificial Intelligence Laboratory (CSAIL).

MS Computer Science, Stanford University

September 2021 — June 2022

BS (Honors) MCS, Stanford University

September 2018 — June 2022

Mathematical and Computational Science (MCS) is an interdisciplinary quantitative major co-organized by computer science, management science & engineering, statistics, and math departments. I was selected as one of four student major advisors during my senior year.

Links

[Portfolio](#) (for more information and reflections) [Youtube](#) (for software project demos)

Employment History

Graduate Research Assistant, MIT CSAIL, Cambridge

June 2023 — Present

I am a driver of multiple projects within the MIT-IBM Watson AI lab, working with client companies on graph-based design and generative AI for scientific discovery in future industries.

- Advisor: Wojciech Matusik
- I work on multiple projects within the MIT-IBM Watson AI lab (PI: Jie Chen), working with client companies on graph-based design and generative AI for scientific discovery in future industries
- I also collaborate with domain experts from other MIT departments to simultaneously develop new ML methods and close the loop experimentally
- I actively brainstorm ways to 1) combine symbolic and neural descriptions for graph-based design, and 2) incorporate Large Language Models within the loop of scientific discovery

Mid-level Machine Learning Engineer @ TikTok (Search Team) , Mountain View

August 2022 — September 2022

- Owned Live (one of main verticals: video, music, users, live, etc.) search
- Created AB tests and conducted human/offline evaluation on worldwide users for new live-related strategies, including new features, updated CTR/playtime/etc. ML models

Full-time Researcher @ Stanford SLAC , Palo Alto

June 2022 — September 2022

- Drove a large-scale simulation research project by combining neural solvers and first-principles solvers to model the multi-scale dynamics of laser-plasma ion acceleration, working with data of first-principles simulations of laser-plasma interaction
- Collaborated closely with physicists to translate scientific hypotheses into experimental ablations or tests, learned about latest research in high-energy density physics, and shared findings with rest of the department

Co-founder, Tech Lead @ Demodraft

July 2020 — September 2020

- Led a team of part-time volunteers to deploy our Beta within 1.5 months (subscribe on demodraft.org to get access to the Beta!)
- Got into Berkeley SkyDeck, **one of the nation's top college accelerators (<10% acceptance)**

Research

GNN for learning NP-hard problem (Graph ML)

January 2023 — September 2023

- Author order: Michael Sun (solo author)
- Conceptualized a novel self-supervised way to learn a classical graph NP-hard problem, achieving near-optimal solutions in 1% of the time as integer programming
- Demonstrates representational transfer to downstream position-aware graph ML tasks
- Published in Neurocomputing ([link](#))

ML for high-energy density plasma simulation (ML for simulating physics)

January 2022 — Present

- Author order: Tailin Wu, Michael Sun, Jason Chou, Pranay Samala, Sithipont Cholsaipant, Sophia Kivelson, Jacqueline Yau, Rex Ying, Paulo Alves, Jure Leskovec (Professor), Frederico Fiua (PI)
- Collaborated with physicists at Stanford's Linear Accelerator Laboratory (SLAC) to use ML to accelerate the simulation of laser-driven ion acceleration using a neural-hybrid solver, having weekly meetings with both domain experts and project members to foster understanding
- Implemented the entire pipeline, integrating together particle-in-cell components and neural components to model high-intensity laser-plasma interactions for training, and wrote custom scripts for stable long-term self-consistent rollout with analysis, visualization, debugging tools

- Formulated final experiments demonstrating our system can faithfully model the laser-plasma multi-scale dynamics while maintaining a good tradeoff between speed and accuracy
- Publication status: NeurIPS AI4Science workshop ([poster](#)), Nature MI submission ([preprint](#))

Privacy aware GNN-based recommender system for cold start (RecSys, NLP)

December 2020 — June 2021

- Author order: Michael Sun, Andrew Wang
- Introduced a novel representation that avoids privacy concerns of prior user-profiling approaches but is expressive enough for bootstrapping cold start recommendations
- Constructed a new dataset for anime recommendations, AnimeULike, a heterogeneous graph of 130k+ interactions between ~13k users and 10k animes from popular Anime website myanimelist.net, containing textual reviews, item-item recommendations, synopses, etc. released dataset to public ([link](#))
- Designed and implemented DeepNaniNet recommendation framework for cold start, designing rich graph representation space using GNN with language encoder
- Formulated and executed experiments demonstrating significant performance improvement over previous cold start solutions
- Published in Springer IJDSA special issue on next-generation recommender systems ([link](#))

Continual representation pretraining (CV, NLP)

March 2021 — Present

- Author order: Michael Sun*, Ananya Kumar*, Divyam Madaan, Percy Liang (Professor)
- Devised and executed the project end-to-end, introducing a simple yet effective technique from transfer learning that can be used as a standalone method to improve fine-tuning in continual representation learning setting without increasing memory cost, but can also be coupled with continual learning (CL) algorithms like SI, DER, LwF to improve performance
- Formulate a technique variant which achieves SOTA accuracy on a NLP BERT-based CL baseline
- Publication status: CVPR CLVision workshop ([preprint](#))

Zero-shot auto-annotation to leverage unlabeled data (NLP)

September 2021 — Present

- Supervisor: Professor Monica Lam, worked with: Mehrad Moradshahi, author order TBD
- Built an automated error analysis pipeline to find typical error patterns made by a zero-shot model trained on translated data
- Synthesized dataset constructing negative examples with large coverage via perturbations to the dialogue state statistically consistent with analysis of auto-error-analysis
- Train error classifiers using synthesized dataset and demonstrate error detection in-the-loop can significantly improve zero-shot performance
- Published in ACL ([link](#))

ML for teaching an industrial-leading robot arm dexterous manipulation (RL for robotics)

June 2020 — September 2020

- Developed deep RL algorithms on Think Tank Team's flagship prototype and cooking assistant concept, Samsung Botchef
- Brought in OpenAI's latest research on applying the PPO algorithm to train actor-critic policies in an open-ended physics environment which mimics a real kitchen, using software/tools like ray, tune, MuJuCo and importing geometric meshes and robot URDF
- One key contribution: replaced the team's prior approach of inverse kinematics, or basic-RL approach precomputing the waypoints, by designing gym environment that encourages open exploration and planning its own trajectories which mimic human-like behavior

- Iterated on experiments to identify key hyperparameters, implementing Bayesian search algorithms, and researched sim2real transfer for asymmetric actor-critic
- Delivered results that were a POC and went on to being demo'ed at CES. Videos of the POC can be found on [portfolio](#)
- Publication status: confidential R&D work, submitted for patent

ML for seizure detection with a neurologist (CV for medicine)

January 2020 — April 2020

- Supervisor: Daniel Rubin, worked with: Chris Lee-Messer
- Augmented seizure detection models with unlabeled patient video timestamped to the world's largest EEG from Stanford hospital
- Devise unsupervised patient tracking techniques (using SiamFC)
- Used tracker as a filter to find moments when the patient is present (eliminating false positives from the EEG model's outputs), and conceived it as an input source into the multi-modal cross-attention EEG model

Internships

MS Researcher @ Stanford Network Analysis Project and Stanford Artificial Intelligence Laboratory, Stanford Open Virtual Assistant Laboratory

March 2021 — June 2023

- See my website for descriptions of the projects and my involvement levels.

Machine Learning Intern @ Spotify

June 2021 — July 2021

- One of 3 inaugural interns in Spotify's ML Platform team
- Began Bayes optimization feature in ML Home, Spotify's productivity hub for all Spotify's applied ML teams
- Proposed pipeline to identify causes for underperforming user segments, used Kubeflow Pipelines for end-to-end ML workflows

Machine Learning Research Intern @ Samsung Research America, Think Tank Team (SRA, TTT), Mountain View

June 2020 — September 2020

- Delivered proof-of-concept: training the SARAM Botchef, TTT's flagship robot arm, to stir kitchen pots in simulation
- Designed and built a goal-based **OpenAI gym** simulation environment, using MuJoCo physics engine, working with geometric meshes and robot URDF
- Iterated >100 experiments with the PPO algorithm, a gradient-based actor-critic approach using ray and tune (RL libraries)
- Designed and implemented techniques to encourage exploration and sim2real transfer

BS Researcher @ Stanford Quantitative Imaging and AI Lab

January 2020 — April 2020

- See my website for descriptions of the projects and my involvement levels.

Machine Learning Intern @ Synaptics , San Jose

July 2019 — September 2019

- Designed and iterated over hundreds of experiments to optimize object detection models and **deployed a real-time video logo detection model** using the Single Shot Multibox Detector (SSD) for on-device real-time-fps inference

- Gained valuable machine learning intuition/expertise by balancing the latest ideas from research papers and trial-and-error
- Deployed an end-to-end inferencing pipeline for **real-time detection** into an Android APK
- Experimented in **Keras**, **TensorFlow** and converted to **TFLite** ops with attention to performance-speed tradeoffs
- **Demo'ed at International Broadcast Conference (IBC) 2019** on Synaptic's new SOC (VSR 371)
- See demo clip and slide deck I presented to the CTO team [here](#)

Software Projects

Selected Examples (demos & details in portfolio)

April 2018 — Present

- OtakuRoll.net (Website) (12/20-present): World's only NLP-driven anime recommendation website that serves guest users, built on DeepNaniNet GNN-based cold-start system
 - Tech stack: **Vue.js**, **Django REST**, **PyTorch**, **Firebase**, **AWS ELB**, **EC2**, **S3**, **Route 53**
 - World's only NLP-driven anime recommendation website that serves guest users, built on DeepNaniNet GNN-based cold-start system
- demodraft.com (Web app) (7/20-9/20):
 - **Django REST + Vue.js** full stack web platform using **AWS EC2**, **S3**, **RDS** and **Route 53**
- Quizkly.com (Web App) (10/18-10/19):
 - **Full-stack NLP application** that takes in corpuses of text and generates MCQ quizzes with blanks (see portfolio for demo)
 - **Full-stack MVC app with the Django REST, React, AWS**
 - Built as a member of Pear Garage, a VC accelerator program, as **one of two dozen selected out of hundreds** of Stanford applicants
- aiRoute (iOS App Store) (6/18-9/18):
 - An end-to-end running app built with **Swift** that auto-generates running routes
 - Embedded map views, live navigation, and voice controller SDKs
 - Deployed **RESTful API** in **Flask** hosted on **Heroku** with **Firebase NoSQL**
- aiFood (iOS App Store) (4/18-11/18):
 - An end-to-end meal generation app built with **Swift** that automates macro-counting
 - Generates ingredient lists that precisely satisfies nutritional needs

Skills

C/C++, Python, JavaScript, Swift, Java

ML: TensorFlow, Keras, TFLite, PyTorch, DGL

Frameworks: Django (+REST), Flask, Vue.js, ReactJS, React Native, iOS native

Services: AWS EC2, RDS, S3, Firebase, Kubeflow, Heroku, Git, iOS App Store

Paradigms: Client-server, MVC, MapReduce

Major Academic Achievements

Intel International Science and Engineering Fair (ISEF)

October 2016 — April 2017

Second Place Grand Award, Category of Mathematics, at World's largest science/engineering exhibition, project titled "Bounds on metric dimension for families of planar graphs" ([arXiv link](#))

United States of America Mathematical Olympiad (USAMO)

May 2017

Ranked *top 120* on USA's most prestigious math competition (~250/50000 qualify each year), scored *top 30 nationwide among ~1,000* on AIME, Distinguished High Honor roll on AMC 10 & 12