Marco Yang

510-738-8001 myang2@caltech.edu linkedin/marco-yang-17189a203 github/mcrco marcoya.ng

EDUCATION

California Institute of Technology *Pasadena, CA* Dec 2026

Computer Science GPA: 3.91

SKILLS

Languages C, C++, Java, Golang, Python, Typescript

Frameworks/Architectures Pytorch, Tensorflow, REST, Langchain

Misc. Linux, Shell, Git, Computer Vision, Generative AI

EXPERIENCE

Software Engineer Intern | *Glean* Jun 2025 - Sep 2025

- AI team (LLM Assistant + Agent Quality) at late-stage startup that provides Search and Assistant for enterprise data
- Implemented, deployed, and monitored LLM-as-a-judge evaluations and dataset pipelines for agents.
- Pushed new automated prompt optimization feature for Glean agent builder.
- Miscellaneous features and bug fixes in Glean's LLM workflow engine/agent backend framework in Golang.

ML Researcher | *Caltech Vision Lab (Perona Lab)* Jun 2024 - Dec 2024

- Researched novel approach to cross-domain video action recognition utilizing generative diffusion models
- Injected into CompVis' Stable Diffusion 1.5 and Hugging Face implementation of Modelscope (SOTA text-to-video model)
- Implemented multiple classification heads (MLP, 3D CNN, Transformer) for action classification on diffusion features
- Used PyTorch, WandB, Huggingface
- No publication, but learned a lot about machine learning and observed random cool stuff like transformer attention maps in videos

Full Stack Software Developer Intern | *Fathomd* Aug 2021 - Jun 2023 (Seasonal)

- Developed educational video games for business schools like MIT Sloan, with a focus on the instructor dashboard.
- Performed full-stack development, including creating React components, adding routes for new pages in the frontend using Meteor (Ironrouter), developing REST API endpoints, and modifying DTO/DAO layers in a Dropwizard Java backend.
- Implemented a redesigned instructor dashboard with new "classes" feature, enabling instructors to manage tailored game sessions for different classes.

PROJECTS

Caltech RAG | **github** | **link**

- Hated Caltech's own catalog and course review website so I made an LLM with retrieval-augmented generation (RAG) app that indexes Caltech's course reviews and catalog
- Scraped and embedded two years of internal course reviews as well as latest course catalog into vector database
- Hybrid search using both vector embedding cosine similarity as well as lexical search (BM25)
- Built with Flask and Langchain/Langgraph

Astro Duel Clone

- Clone of the local multi-player video game Astro Duel written in C and and compiled to web assembly for browsers
- Implemented asset cache, physics engine for mass, bodies, collision handling, momentum, force, and torque, graphics using SDL, and bot players

RELEVANT COURSEWORK

Data Structures, Algorithsm, Advanced Machine Learning, Computing Systems, Applied Linear Algebra, Statistical Inference, Information Theory

HONORS & ACCOLADES

IMC Prosperity Trading Competition | *Top 3 in 2 rounds* April 2025

- Placed 2nd out of 13,000 teams in Round 2 and 3rd out of 13,000 teams in Round 3

William Lowell Putnam Mathematical Competition | *Top 1000 (scored 19 points)* December 2023

USA Computing Olympiad | *Gold Division* February 2021

American Math Competition/American Invitational Math Exam | *3x AIME Qualifier* November 2021