YUHENG DING

706-238-1952 • AaXDing.github.io • linkedin.com/in/yuheng-ding

EDUCATION

Carnegie Mellon University Expected 12/2026

MS, Intelligent Information Systems

University of California - Los Angeles 09/2021 – 06/2025

BS, Computer Science & Applied Mathematics, GPA: 3.88 / 4.0

• Courses: Algorithm and Complexity, Computer Network, Machine Learning, Database Systems, Data Structure, Computer Organizations, Deep Learning, Computer Vision, Natural Language Processing, Graph Theory, Linear Algebra.

TECHNICAL SKILLS

Java (Tomcat, Spring Boot), **Python** (Django, PyTorch, NumPy), **C++**, **SQL** (MySQL, PostgreSQL, SQLite), **Redis**, **MongoDB**, **JavaScript** (React), **AWS**, **Azure**, **Google Cloud**, **Git**, **Bash**, **Linux**, **Docker**, **Kubernetes**, **HTML**, **CSS**

PROFESSIONAL EXPERIENCE

VisionX | Full-Stack Software Engineer Intern 07/2023 - 10/2023

- Collaborated with a team of four engineers to architect a full-stack e-commerce platform, with a focus on promoting a healthy lifestyle and fostering environmental sustainability.
- Engineered a product page UI template using Python Django framework for backend and MongoDB for data persistence; streamlined automatic page creation from database; enhanced scalability and amplified product creation efficiency by 400% over manual page inputting.
- Implemented user information gathering, payment workflow with Stripe API, and email order confirmations; achieved a 30% reduction in checkout process response time.
- Collaborated with a colleague to enhance online consulting session reservation system, integrating voice and video chat functionalities by using WebSocket and WebRTC technologies, resulting in 60% decrease in reservation confirmation time.

RESEARCH EXPERIENCE

UCLA-NLP Mentor: Kai-Wei Chang, Nanyun Peng | Undergrad Researcher 06/2024 - Present

- Developed a comprehensive benchmark for evaluating visual language models (VLMs) on perception and reasoning quality in QA tasks, with testing on over 25 VLMs to assess self-critique and correction abilities. (CVPR 25' Accepted) Link
- Developed an intuitive annotation tool that improved the annotator experience, enabling efficient curation of over 1,600 fine-grained annotated samples for a high-quality dataset.
- Introduced a novel evaluation metric, namely VISCore, which incorporated answer accuracy, step-wise critique accuracy, and critique quality, leading to more accurate assessments.
- Implemented an agent-based critique generation method at each reasoning step, achieving an 8% increase in critique accuracy.

UCLA ARNI Lab Mentor: Christina Fragouli | Research Intern 07/2023 - 09/2023

- Refined a graph theory solution to minimize tests and time used in identifying infection within a community network using Python.
- Developed data pre-generation script using Python, resulting in a 70% reduction in test time for a population size of 20k.
- Improved testing script with Jupyter and JSON that enabled test creation and logging with 20% overhead reduction.
- Originated a community-aware diagonal splitting algorithm using a divide-and-conquer strategy which factored in community structure; the new algorithm needed 10% less tests compared to 5 other commonly-used algorithms.

PROJECT EXPERIENCE

High Performance URL Shortening Web Server | Tech Lead & Individual Contributor 04/2025 - 06/2025

- Led a team of 4 to design and build a low-cost, high-performance asynchronous TCP server in **C++** with **Boost.Asio** using test-driven development; authored a detailed PRD and system design document to guide architecture and implementation.
- Designed and implemented a scalable URL shortening service with **Redis** caching and **PostgreSQL** persistence, leveraging connection pooling to support 1,000+ concurrent users.
- Deployed the service on **Google Cloud Platform** using **Docker** and Compute Engine; configured CI/CD with **Cloud Build**, including automated testing, health checks, and container recovery for production reliability.

LLM Evaluation | Team Leader 01/2024 - 03/2024

- Led a team of 4 in researching prompting techniques, enhancing the factual accuracy and fairness assessment of LLMs on textual claims.
- Developed a zero-shot evaluation framework using Microsoft Phi-2 and Mistral-7B, incorporating GPT-generated evidence and chain-of-thought reasoning to assess designated tasks, achieving a 76% accuracy rate

Text Guided Image Editing with Diffusion | Undergrad Researcher 02/2024 - 03/2024

- Collaborated in a team of 4 to pioneer image editing solely based on target prompts using DiffEdit with Bootstrapping Language-Image Pre-training (BLIP) and Stable Diffusion Model; significantly reduced prompt restrictions and boosted the efficiency of the image creation process.

Job-Search | Independent Developer 10/2023 - 01/2024

- Constructed a job search platform that personalized job recommendations merely based on user views and likes; incorporated Search, Favorite, and Recommendation functionalities in **Java Servlets** with **REST APIs** to efficiently handle HTTP requests and responses.
- Implemented frontend UI components using **JavaScript**, **HTML**, **CSS**, and **Ajax**, facilitating user login and enabling users to efficiently search and apply for positions.
- Designed AI-based keyword extraction algorithms and history-based recommendation to implement a job recommendation engine; Prioritized user safety through session-based authentication.
- Configured and deployed on **AWS EC2** virtual machine, combining **Redis** and **Amazon RDS** to cache and persist search and favorite related data; scaled the system to be able to serve 10k queries per second (QPS) and reduced latency by 90%.

Bruin-O-Bruin Web App | Project Lead 10/2022 - 12/2022

- Coordinated with a team of 5 to build a full-stack elimination web game using Node.js, React framework, and SQLite.
- Led implementation of core features including game randomization, discussion forum, and scoreboard; ensured instantaneous user interactions.
- Launched account management features, including user registration, password protection with hashing, and password reset by using Node.js and SQLite; the feature led to 30% decrease in user support requests.
- Launched on Azure VM and shared the app to our class forum; achieved over 50% class participation rate where players could compete on score leaderboard.