MATTHEW SHEN

matthewrshen@gmail.com | linkedin.com/in/matthewrshen | github.com/Sariel1563

EDUCATION

UNIVERSITY OF CALIFORNIA, SAN DIEGO — San Diego, CA
B.S. Computer Science w/ Minor Business Economics — Sep 2020 – Jun 2024

- *Relevant Courses:* Deep Learning, Natural Language Processing, Machine Learning, Recommender Systems & Web Mining, Advanced Data Structures, Object-Oriented Design, Linux & Vim Techniques, Database Systems
- *Technical Skills:* C++, C, Java, Python, PyTorch, TensorFlow, DSPy, Arm, Shell, SQL, JavaScript, R, Linux, OpenAI Models, Lamini, Hugging Face API, LLM, LangChain, CoreNLP, Gradio, Search Engine Optimization
- *Interests*: table tennis, skiing, rock climbing/bouldering, chess, badminton, hiking, exploring new places

PROFESSIONAL EXPERIENCE

INTEL: Technical Engineer Intern — Sep 2023 – Jun 2024

- Managed 898 Windows servers through ILO configuration with DNS, installing OS, and settling incident tickets
  - Rebuilt operating systems on 40+ HP and Supermicro servers through Cobbler or by mounting virtual ISO
  - Produced 10+ statistical diagrams in Power BI for DevOps team presentation by filtering databases with SQL
- Made Python/Powershell scripts to automate repair tickets, OS installs, setting of ILO name/IP & subnet masks
- Collaborated on 4 Ansible Playbooks to improve infrastructure as code and fully automate software installations

MEMVERGE: Software Engineer Intern — Jun 2022 – Aug 2022

- Created Linux kernel virtual machine with Compute Express Link enabled to emulate CXL persistent memory
- Tested QEMU emulation of CXL DRAM device resulting in non-compatibility and a switch to physical cards
- Troubleshooted and recorded memory of Samsung and Montage CXL cards and compared with non-cxl servers
  - Using Intel MLC, discovered 50% decrease in latency and 1500% increase in bandwidth between servers
  - Worked with hugepages to configure the daemon I created, allowing me to do tests on a memory machine
  - Performed redis tests on efficacy of memory machines with varying DRAM tiering, resulted in 18% decrease in latency with results presented by CEO at Flash Memory Summit
- Updated memory tiering test to allow multithread execution as well as specific cache workload selection

CLOSE THE GAP FOUNDATION: Technical Marketing Intern — May 2021 – Aug 2021

- Refactored and optimized the Google Ads Campaign and created 4 new webpages using Squarespace and Ghost
  - Created SEO framework tracking 10+ user statistics across all webpages and maintained full website solo
  - Increased web traffic to under-viewed webpages such as Achievement Gap and FGLI Communities by 400%

PROJECTS

AI RESUME SCREENER AGENT — Jun 2024 – Present

- Built AI agent using CrewAI to give feedback on resume and portfolio strength and compatibility with target jobs
  - Used GPT 3.5 and Serper to create agent crew to research job requirements and rewrite applicant profile
  - Trained model on Nvidia GPU and Cuda platform with maximum memory usage nearing 6 GB

AI CHATBOT w/ RAG ENHANCEMENT — Feb 2024 – Mar 2024

- Created interactive AI chatbot using ChatGPT API and improved accuracy by 5% for domain related knowledge
  - Enhanced GPT 3.5 model by employing DSPy framework with RAG to reduce hallucinated answers by 10%

CAMAURAL APP — Aug 2023 – Nov 2023

- Published AI application that recognizes voice command inputs and alters features of a phone's camera
  - Used Google Cloud Speech-to-Text, coreNLP, and Android's Camera API to create the multimodal software
  - Built the capability of recognizing synonymous sentence inputs on the foundation of a graph data structure

AI IMAGE GENERATION APP — Jul 2023 – Aug 2023

- Created app with image recognition/captioning and image generation from descriptions using stable diffusion
  - Used Hugging Face API and Gradio to create UI that correctly identifies image inputs with 97% accuracy

ZOOSEEKER ANDROID APP — Mar 2022 – Jun 2022

- Created route planning app for San Diego Zoo with live location tracking and dynamic route changing capabilities
  - Incorporated advanced data structures like mock objects, adapters, and observers to link 30+ classes together
  - Obtained latitude and longitude coords of zoo exhibits by parsing a database with Room and MySQL queries
  - Used Google Location API and Dijkstra's algorithm to create the shortest path to visit all selected exhibits