

JACKSON NEWMAN

Redwood City, CA, 94065 | 650-649-8204 | jnewman167@gmail.com | [linkedin.com/in/jacknewman](https://www.linkedin.com/in/jacknewman) | <https://github.com/JNewman-cell/> | jacksonnewman.netlify.app/

WORK EXPERIENCE

Majestic Labs AI

Los Altos, CA

Software Engineer

September 2024 - October 2024

- Configured and deployed 5 AWS and GCP Linux VMs for AI model performance testing and benchmarking.
- Benchmark-tested 10 AI models in production using inference engines like vLLM and TensorRT-LLM.
- Improved Stable Diffusion performance with NNCF quantization, cutting image generation time by 80%.
- Deployed an end-to-end Flask web app with Nginx to showcase OpenAI-like LLM inference via HTTP.
- Developed a Flask API server capable of handling hundreds of image generation requests.
- Automated batch size testing for various LLMs using Llama.cpp, reducing data collection time by 30%.

AMD

San Jose, CA

Software Engineer Intern, Internal Development Tooling

June 2023 - September 2023

- Accelerated Vivado constraint processing by 50% with a new C++ pattern matching function.
- Developed unit tests achieving 100% coverage to ensure performance and accuracy for Wildcard Matching.
- Reduced Vivado memory usage by 2% by refactoring code to utilize Tessil C++ Hash map package.
- Decreased testing time by 80% through automation of memory and encryption performance tests.
- Cut Vivado project update time by 50% by automating encryption key upgrades for custom IP.

Shellie.us

San Francisco, CA

Full Stack Software Engineer Intern, MERN Stack

June 2022 - September 2022

- Created an edit page for online exhibits, allowing 50+ admins to update marketing and authorized users.
- Enhanced user experience, eliminating 100% of 30+ monthly inquiries about exhibit updates.
- Streamlined new exhibit page launches, reducing time by 60% (from 5 days to 2 days).

PROJECTS

TinyGPT | *Python, PyTorch, Numpy, Self Attention, Transformers*

- Implemented a cutting-edge language model using PyTorch, enabling the generation of text from input.
- Utilized advanced NLP techniques inspired by GPT-2, including Transformers, for effective pattern capture.
- Demonstrated expertise in training and optimization within the GPT-2 framework for text generation.

Autoencoder Visual Classification | *PyTorch, Numpy, Seaborn, Pandas, Matplotlib*

- Trained autoencoder for precise letter recognition and transistor positioning on MNIST and MVTEC dataset.
- Engineered efficient autoencoder architecture with optimized feature extraction for image classification.
- Demonstrated 100% accurate image classification using threshold pixel-by-pixel differences.

Wafer Map Failure Classification | *PyTorch, Numpy, Seaborn*

- Architected a 14-layer model classifying wafer types with > 90% accuracy using the WM-811k dataset.
- Devised Augmentation layers for a noise-resistant network, slashing 80% of random chip failures.
- Analyzed and visualized accuracy and loss throughout iterations to identify over- or under-training.

EDUCATION

University of California, Santa Barbara

Santa Barbara, CA

Bachelor of Science in Computer Engineering

September 2020 - June 2024

3.7 GPA, Dean's Honors, Relevant Coursework:

Data Structures, Algorithms, Operating Systems, Applied Machine Learning and AI, Embedded Systems

TECHNICAL SKILLS

Programming Languages: C++, Java, Python, C, SQL (Postgres), JavaScript, HTML/CSS

Frameworks: React, Node.js, Flask, Bootstrap, Material-UI, Storybook

Developer Tools: Git, Docker, Jira, Google Cloud Platform, Confluence, Perforce, NPM, Firebase, GitHub Actions

Libraries: pandas, NumPy, TensorFlow, PyTorch, Keras, Seaborn, matplotlib