JONATHAN GUIANG

jkguiang@gmail.com
+1-858-880-5819

jguiang.com | github.com/jkguiang | linkedin.com/in/jonathanguiang

Languages · Python, C++, Javascript, SQL, Bash, HTML/CSS · **Packages** · PyTorch, Tensorflow, Keras, XGBoost, pandas, Numpy, Scikit-learn, Flask · **Tools** · Docker, Kubernetes, Git, VIM, CUDA, React

EXPERIENCE

Data Scientist developing novel AI products that support the wellbeing of millions of people living with diabetes and empower them in managing their condition. Physics PhD with a background in studying the fabric of existence and beginning of time with petabytes of data. Passionate mentor and aspiring leader.

Senior Data Scientist/ML Engineer, Dexcom *Jul. 2024 - Present*

- Led the development of a multimodal LLM (image classification + content generation) features for the companion app to Dexcom's glucose monitors. Reduced misclassification rate by 50%.
- Designed and implemented API endpoints and MySQL database integrations (> 8k new lines of code) for the backend of the internal mobile app used to test new LLM-based features.
- Developed a multithreaded pipeline for assessing the performance of LLMs and curated a large multimodal dataset for it (10k examples). Championed the evaluation of LLMs at scale.
- Established our team's coding standards and implemented automated documentation systems.
- Conducted interviews, mentored junior colleagues, and contributed to code reviews.

Graduate Researcher, UC San Diego *Mar. 2020 - Jun. 2024*

- Led the training of a deep neural network with a specialized loss function tailored for high-precision particle physics. Achieved a significant 50% reduction in systematic uncertainty for a pivotal Higgs boson measurement.
- Engineered a deep neural network to discern particle trajectories from point clouds (100k points) within a GPU-accelerated C++ algorithm. Resulted in a notable 40% decrease in false positive rate.
- Designed and implemented a networking solution facilitating the global dissemination of exabytes of LHC data. Deployed demonstration and benchmarking testbeds with Kubernetes.
- Architected a robust C++ framework now used at 4 universities for processing petabytes of LHC data across globally distributed computing systems. Supervised a team of 2 junior developers.
- Directed operations for 3 petabytes of physics data hosted at the UCSD "Tier-2" computing facility. Managed the seamless bulk transfer of this dataset to a new infrastructure.

Director, EXPAND Program, UC San Diego *Jan. 2020 - Jun. 2024*

- Co-founded a program that provides underserved undergraduates with research opportunities.
- Grew the program from inception to one of the top-3 funded programs at the UCSD Student Success Center and directed the program coordination team.
- Helped over 40 students secure full-time positions (Raytheon, Thermo Fisher), internships (Apple, NASA JPL), admission to graduate programs (UPenn, UCLA, UCSD), and publications.

Student Developer, Google Summer of Code *May 2019 - Sep. 2019*

- Produced a pipeline in Python that regularly retrieves LHC data cache access records from a database via Spark SQL and computes diagnostic metrics.
- Designed a Grafana dashboard that displays the computed metrics to the engineers that maintain the data caches, allowing them to identify previously invisible problems and inefficiencies.

EDUCATION

University of California, San Diego San Diego, CA
PhD in Physics, MS in Physics *2019 - 2024*

University of California, Santa Barbara Santa Barbara, CA
BS in Physics *2015 - 2019*