# Keyu Long

San Diego, CA | (858)-319-9628 | kelong@ucsd.edu | https://keyu-long.netlify.app | LinkedIn | GitHub

## EDUCATION

**University of California, San Diego** **San Diego, CA**

*MS in Computer Science , GPA: 4.0/4.0* Sep. 2024 – Mar. 2026

- Relevant Coursework: Principles of AI-Probabilistic Reasoning and Learning, Natural Language Processing

**University of California, San Diego** **San Diego, CA**

*BS in Data Science, GPA: 3.9/4.0* Sep. 2019 – Mar. 2024

- Relevant Coursework: Deep Learning, Recommender System and Web Mining, Probabilistic Modeling and ML

## RELEVANT SKILLS

**Computer Science:** Data Structure, Algorithm Design, Object Oriented Design

**Programming:** Python, GO, Java, C, R, Pandas, Git, Kubernetes, Docker, Linux, SQL

**Machine Learning:** PyTorch, Scikit-Learn, CNN, Transformers, CV, NLP, DeepSeek, RAG, MCP

**Language:** Mandarin (Native), English (Professional)

## RELEVANT EXPERIENCE

**Huawei** **Shanghai, China**

***AI Engineer Intern*** Jun. 2025 – Aug. 2025

- Studied key topics in AI infrastructure, including operator fusion, automatic differentiation, computation graphs, distributed parallelism, benchmarking methodologies, and emerging protocols such as MCP
- Participated in an LLM supervised Fine Tuning project tailored for the company's internal programming language to improve workflow automation and coding assistance
- Prepared training data by processing official documentation, code examples, and constructing domain-specific Chain-of-Thought (CoT) reasoning samples
- Adopted reinforcement learning paradigms (inspired by DeepSeek R1) to scale high-quality training data from limited annotations, effectively boosting model performance on internal tasks

**University of California, San Diego** **San Diego, CA**

***Research Assistant for Prof. Garrison W. Cottrell*** Mar. 2023 – Jun. 2024

- Applied divisive normalization in PyTorch to enhance computer vision models, inspired by principles from primate vision, increased performance of shallow CNNs by 20% over ReLU activation.
- Conducted extensive testing of divisive normalization layers on the ImageNet dataset, comprising 1,000 classes and 14M+ images, and against different types of noises.
- Enhanced data loading efficiency by replacing the CPU **data pipeline** with NVIDIA Data Loading Library (DALI), resolving CPU bottlenecks and achieving a **2x** speedup.

## RELEVANT PROJECTS

**Fine-tuning LLAMA 1B with LORA on Toxic Text Classification** [Link] **San Diego, CA**

***Developer*** Nov. 2024 – Dec. 2024

- Utilized and preprocessed the Jigsaw toxic comments dataset for a text classification task.
- Designed prompt for LLM to classify the text in contrast to task specific models like LSA, Word2Vec, and BERT
- Fine-tuned LLAMA 3.2 1B using LoRA with PEFT locally and gain 10% accuracy gain compares to Word2Vec baseline model.

**New Initialization Mechanisms for CNNs** [Link] **San Diego, CA**

***Team Leader*** Oct. 2023 – Mar. 2024

- Developed a new initialization method for Convolutional Neural Networks inspired by advanced research on Average Gradient Outer Product (AGOP) and Neural Feature Matrix (NFM).
- Implemented the method in PyTorch and tested on models like VGG11 across datasets such as SVHN, CIFAR-10, and tiny-ImageNet, achieving a 2%-6% improvement in performance over state-of-the-art Kaiming initialization methods.