

Amogh Patankar

408-597-2478 | apatankar@ucsd.edu | [linkedin.com/in/apatankar22](https://www.linkedin.com/in/apatankar22) | github.com/apatankar22 | apatankar22.github.io

EDUCATION

University of California, San Diego <i>M.S., Computer Science & Engineering; Concentration: AI & Machine Learning</i>	September 2023 – June 2025 <i>La Jolla, CA</i>
University of California, San Diego <i>B.S., Data Science</i>	September 2020 – March 2023 <i>La Jolla, CA</i>

PROFESSIONAL EXPERIENCE

AI Applications Development Intern <i>Advanced Micro Devices (AMD)</i>	September 2024 – December 2024 <i>San Jose, CA</i>
<ul style="list-style-type: none">Developed software to optimize execution of generative AI workloads (Llama-2, Llama-3, Stable Diffusion, etc.) on RyzenAI neural processing unit (NPU).Benchmarked generative AI workloads on Ryzen and competitor hardware (Snapdragon X Elite, Lunar Lake), specifically time to first token, and tokens/sec, i.e. latency and throughput, as metrics.Evaluated NPU latency and utilization during concurrent execution of generative AI workloads and high-fidelity AI effects (Microsoft Studio Effects).	
Data Scientist, Generative AI Intern <i>Marvell Technology Inc.</i>	June 2024 – September 2024 <i>Santa Clara, CA</i>
<ul style="list-style-type: none">Architected an end-to-end ETL pipeline consisting of data engineering, analysis and visualization using numpy, pandas, sklearn, Tableau, etc. Data was ingested using Apache and AWS Kinesis, and stored in SnowflakeDB, and AWS S3.Utilized generative pretrained transformer (GPT) models to generate synthetic data, leveraging parameter efficient fine-tuning (PEFT) techniques such as low-rank adaptation (LoRA).Implemented reinforcement learning (RL) and deep learning methods for DSP parameter optimization.Integrated large language models (LLM) and retrieval augmented generation (RAG) to automate hardware modeling process.	
Researcher <i>Stanford University School of Medicine</i>	June 2023 – August 2024 <i>Palo Alto, CA</i>
<ul style="list-style-type: none">Performed data engineering, analysis, and visualization through various methods in Python and R; specifically, pandas, geopandas, numpy, matplotlib, scikit-learn and ggplot. Developed unique statistical packages composed of chi-squared and Fisher tests.Led research teams mentored by Dr. Gross, Dr. Palaniappan, and Jin Long. <u>Opioid overdose research</u> published in British Journal of Anaesthesia, and diabetes research in preprint at Journal of Asian Health.	
Software Development Engineer Intern <i>Amazon Web Services (AWS)</i>	June 2022 – September 2022 <i>Seattle, WA</i>
<ul style="list-style-type: none">Led architectural changes in Lex ASR (Automatic Speech Recognition) Services and DataHub, improving latency for conversational AI models. Conducted A/B tests to evaluate architectural changes for customer use.Performed cohort analysis to segment critical and non-critical customer data in DataHub, enhancing Lex ASR schemas, and enabled compliant storage of all data using AWS Kinesis, S3, and Lambda.Applied time series analysis on AWS CloudWatch metrics to track and optimize the performance of ASR Service, leading to lower response time and accelerating customer request resolution by up to ~75%.	
Research Intern <i>Scripps Research Translational Institute</i>	June 2021 – August 2021 <i>La Jolla, CA</i>
<ul style="list-style-type: none">Developed an R package to estimate genetic regulatory variation using a confidence interval estimation methods, applying simulation-based techniques like bootstrapping to evaluate the estimates' reliability.Implemented binomial distributions and various other statistical concepts, and used sensitivity analysis to quantify the impact of different assumptions on the genetic data from the Genotype Tissue Expression Project (GTEx).	

SKILLS

- Languages:** Python, Java, R, C++, SQL, JavaScript
- Frameworks and Tools:** PyTorch, Tensorflow, LangChain, Pinecone, Keras, numpy, sklearn, pandas, seaborn, plotly, ggplot, AWS [S3, EC2, SageMaker, Kinesis, Lambda], SnowflakeDB, Tableau, Streamlit, git

PROJECTS

- Capstone Project:** Active Learning with Neural Processes for Epidemiology Modeling
- Prediction of Causes of Patient Readmission using Large Language Model(s) (**OpenAI, Llama-7B, BART-Large**)
- Data Science Interview Tool (**GPT-3**, PyTorch, Python)
- Autonomous Vehicle Trajectory using Deep Learning (**Argoverse 2 Dataset**, PyTorch)