

# AMITASH NANDA

San Diego, CA | (858) 319-5516 | ananda@ucsd.edu | linkedin.com/in/amitash01 | amitashnanda.github.io

## EDUCATION

<b>University of California San Diego</b> , San Diego (CA), USA	Expected December 2026
Ph.D. in Electrical and Computer Engineering (Machine Learning and Data Science)	3.7
Research: Distributed ML, Generative AI, Computer Vision, DNN/LLM Optimization, AI/ML Systems, Applied Data Science	
<b>University of California San Diego</b> , San Diego (CA), USA	September 2021 – June 2023
Master of Science in Electrical and Computer Engineering (Intelligent Systems, Robotics and Control) [ <b>Thesis</b> ]	3.7
<b>Odisha University of Technology and Research</b> , Bhubaneswar (OD), India	August 2015 – May 2019
Bachelor of Technology in Instrumentation and Electronics Engineering	4.0

## TECHNICAL SKILLS

<b>Languages/Databases</b>	Python, C++/CUDA, HTML/CSS, JavaScript, SQL, MySQL, GraphQL, MongoDB, Hadoop, Teradata
<b>Frameworks</b>	PyTorch, TensorFlow, Keras, Theano, NLTK, Fast AI, Spark, ONNX, YOLO, Django, Flask, MPI, Hydra
<b>Libraries/APIs</b>	Scikit-learn, Pandas, NumPy, Dask, SHAP, OpenCV, D3.js, Bokeh, OpenMP, TensorRT, PyTensor, REST
<b>Tools</b>	Anaconda, Jupyter, LaTeX, Nvidia cuDNN, GitHub, Bitbucket, Jenkins, Slurm, Grafana, Adobe Illustrator
<b>Cloud/Technologies</b>	AWS (SageMaker, EC2), GCP, Azure, Docker, Kubernetes, Linux, ETL Tools, Hugging Face, HPC

## EXPERIENCE

<b>Graduate Student Research Intern   NERSC (LBNL)</b> , Berkeley (CA), USA	June 2024 – September 2024
<ul style="list-style-type: none"><li>Expanded the load-balancing algorithms in <b>AMReX</b>, parallelized a brute-force approach, developed a hybrid SFC-Knapsack algorithm, improved SFC bisection strategy using <b>painter's</b> and combined Knapsack using the painter's algorithm.</li></ul>	
<b>Teaching Assistant   CSE, HDSI, ECE (UCSD)</b> , San Diego (CA), USA	March 2024 – Present
<ul style="list-style-type: none"><li>Managed <b>six</b> groups of <b>60</b> students for the Data Science Capstone Project, guiding them through <b>ML model optimization</b>, <b>NLP</b>, <b>Gen-AI</b>, and <b>Graph ML</b>. Assisting as a TA in the Advanced Data Structures course (<b>CSE 100</b>) in the spring quarter.</li><li>Facilitated discussions, office hours, and assessments, supported group projects and provided guidance for code checkpoints.</li></ul>	
<b>Software Engineer Intern   Teradata</b> , San Diego (CA), USA	June 2023 – September 2023
<ul style="list-style-type: none"><li><b>Performance Optimization:</b> Enhanced Teradata's Object File Storage (OFS) system by leveraging Primary Index (PI) for efficient local aggregation and join capabilities, achieving <b>20%</b> query cost reduction. Optimized data scenarios for <b>1B</b> rows.</li><li><b>Framework:</b> Developed an automated framework for data and query generation, simulated good and bad cases for <b>DBQL</b> performance benchmarking. Generated <b>512</b> objects with <b>700M</b> rows (good case) and <b>233</b> objects with <b>48M</b> rows (bad case).</li></ul>	
<b>Research Intern   Teradata</b> , San Diego (CA), USA	June 2022 – September 2022
<ul style="list-style-type: none"><li><b>Performance Optimization:</b> Developed predictive models to forecast platform configuration resource usage, analyzed Telemetry data from the <b>Teradata Telemetry Collection Agent</b> (TCA), and improved operational efficiency.</li><li><b>Framework:</b> Designed and implemented an end-to-end automated pipeline, integrated data retrieval from TCA using <b>REST</b> API, performed data-preprocessing, built predictive models, and streamlined platform monitoring and forecasting.</li></ul>	
<b>Graduate Student Researcher   Boolean Lab</b> , San Diego (CA), USA	September 2021 – Present
<ul style="list-style-type: none"><li><b>Computational Biology:</b> Analyzing large-scale biological datasets to identify <b>Boolean</b> relationships between genes and developing computational methods to accelerate the use of AI in pathology and drug discovery.</li><li><b>AI Integration and Model Optimization:</b> Developing multimodal approaches, foundational models, and fine-tuning strategies for vision and language-related task to generate biological hypotheses and identifying biomarkers.</li></ul>	
<b>Software Engineer II   Accenture</b> , Bengaluru (KA), India	September 2020 – August 2021
<ul style="list-style-type: none"><li><b>Application Development:</b> Designed and developed a novel robot software testing application (<b>chaosRobo</b>) utilizing chaos engineering principles to simulate real-world scenarios. Led the <b>first-phase</b> prototype development team.</li><li><b>Automated Framework:</b> Contributed to backend <b>GUI</b> design, integrated <b>Gazebo-ROS</b> functionalities, and successfully deployed the system on <b>AWS S3</b> and <b>RoboMaker</b>, enhancing robot simulation and testing efficiency.</li><li><b>Performance Optimization:</b> Implemented machine learning-based analytics in chaosRobo to assess system robustness, reducing failure detection time and improving overall testing accuracy.</li></ul>	
<b>Software Engineer I   Accenture</b> , Bengaluru (KA), India	July 2019 – September 2020
<ul style="list-style-type: none"><li><b>Bedrock Automation Asset Development:</b> Automated end-to-end user stories for <b>Airbus</b> PPR scenarios using image-based and text-based recognition. Developed a new testing methodology and integrated it into the <b>3D Experience</b> software.</li><li><b>Automated Framework:</b> Worked on customization and integration using <b>RESTful</b> web services and hybrid app development using <b>Apache Cordova</b>, enhanced cross-platform functionality and user experience.</li></ul>	

## PROJECTS

### Dynamic Load Balancing Algorithms in AMReX | NERSC (LBNL) [Link]

June 2024 – March 2025

- Developed a novel hybrid load-balancing algorithm combining SFC and Knapsack, improved SFC bisection and Knapsack using **painter's** algorithm, optimized dynamic load-balancing in AMReX framework on NERSC Perlmutter supercomputer.
- Statistically analyzed the efficiency and runtime of these algorithms for sizes up to **512** ranks. Knapsack combined with painter's achieved highest efficiency ranging **99.8-99.9%**, while SFC with painter's algorithm demonstrated faster runtime.

### CHAI-KTQ: Framework for Scalable LLMs & Efficient Inference | Boolean Lab [Link]

August 2024 – Present

- CHAI Quant employs mixed-precision quantization for clustered attention heads, reducing Key-Value (K, V) cache size by up to **55%** and improving latency by **40%**, all while keeping accuracy deviations below **1%**.
- CHAI Target focuses on targeted fine-tuning of sensitive layers identified through attention sensitivity analysis, ensuring robust predictions and reducing uncertainty in critical tasks.
- CHAI KD enables efficient knowledge transfer from large teacher to lightweight student models, achieving speed gains of **3000** inferences/sec for **125M** models with competitive performance on knowledge-intensive tasks like **PIQA** and **RTE**.

### CPTQuant: Mixed Precision Post-Training Quantization for LLMs | Boolean Lab [Link]

May 2024 – Present

- **CMPQ** adapts the precision level based on canonical correlation analysis of different layers. **PMPQ** optimizes precision layer-wise based on their sensitivity to sparsity. **TDMQ** modifies precision using Taylor decomposition to assess each layer's sensitivity to input perturbation.
- **CPTQuant** achieves up to **4x** compression and a **2x-fold** increase in efficiency compared to Hugging Face **FP16**.
- **PMPQ** demonstrates an **11%** higher compression ratio than other methods for classification tasks, while **TDMQ** achieves a **30%** greater compression ratio for language modeling tasks.

### Communication Efficient Asynchronous Peer-to-Peer Federated LLMs | Boolean Lab [Link]

December 2023 – April 2024

- Developed a secure, efficient, and privacy-preserving federated learning framework in a decentralized setting, achieving **5x** lower latency and **13%** higher accuracy in serverless environments.
- Observed a **76%** reduction in information passing time for asynchronous methods, while PageRank proved most effective at removing anomalous nodes, boosting the global federated LLM model's performance.

### OrgaTuring: Accelerating Organoid Discovery with Vision-AI | Boolean Lab [Link]

May 2023 – Present

- Developed an interpretable DNN model for localization, tracking, and classification of 2D/3D Crohn's disease organoid images with focal loss, G-mean thresholding, and domain adaptation (**75%** accuracy, **0.67** AUC-ROC).
- Exploring zero/few-shot learning and patch-based probabilistic classification, also developing a conformal prediction method using gradient-based sensitivity maps for small sample sizes.

### Expression Gradient of Cancer Suppressor Gene using Vision-AI | Boolean Lab [Link]

May 2022 – May 2023

- Developed a colon gland instance segmentation model using Mask R-CNN and Yolo-v8 to diagnose colon cancer. Mask R-CNN achieved F1/IoU scores of **0.63/0.46** (glands) and **0.59/0.42** (crypts). Annotated first public U-shaped colon dataset.
- Demonstrated that Yolo-v8 outperformed Mask R-CNN with mAP50 scores of **0.937** (glands) and **0.748** (crypts). Predicted **5,000+** U-shaped glands across **25** slides, identifying differentially expressed genes along crypts.

## PUBLICATIONS

- **Amitash Nanda**, Md Kamal Hossain Chowdhury, Hannah Ross, and Kevin Gott, "Exploring Dynamic Load Balancing Algorithms for Block-Structured Mesh-and-Particle Simulations in AMReX," **accepted at ACM PEARC 2025**.
- **Amitash Nanda**, Sree Bhargavi Balija, and Debashis Sahoo, "CHAI-KTQ: A Novel Framework for Scalable LLM and Efficient Inference," manuscript in preparation, 2025.
- **Amitash Nanda**, H. M. Zabir Haque, and Debashis Sahoo, "Leveraging High-Performance Computing for Spatial Transcriptomic Identification of CDX2 Genes in Intestinal Crypts using DNN," manuscript in preparation, 2025.
- **Nanda, A.**, Balija, B., and Sahoo, D., "CPTQuant – A Novel Mixed Precision Post-Training Quantization Techniques for Large Language Models," preprint arXiv:2412.03599, December 2024. [Link]
- Balija, B., **Nanda, A.**, and Sahoo, D., "Building Communication Efficient Asynchronous Peer-to-Peer Federated LLMs with Blockchain," published in Proceedings of AAAI Spring Symposium Series, May 2024. [Link]
- **A. Nanda**, D. Dang, and D. Sahoo, "OrgaTuring: A Novel Deep Learning Framework to Classify Healthy vs. Diseased Colon Organoids," ready for submission, 2025.
- Behroozikhah, M., **Nanda, A.**, et al., "Expression Gradient of Cancer Suppressor Gene Found in Colon Crypt Using Vision-AI," ready for submission, 2025.
- **A. Nanda**, D. Sahoo, and B. Lin, "Novel Vision-AI Techniques for Morphological Discovery in System Biology," Master Thesis, Open Access Publication, UCSD, May 2023. [Link]
- D. Dang, **A. Nanda**, D. Sahoo, and B. Lin, "NeuCASL: From Logic Design to System Simulation of Neuromorphic Engines," Poster Presentation, Formal Methods in Computer-Aided Design, Yale, 2021. [Link]
- R. Vedula, **A. Nanda**, et al., "Computer Vision Assisted Autonomous Intra-Row Weeder," ICIT (IEEE), 2018. [Link]