# CHINMAY DATTANAND KUCHINAD

📞 +1(213) 378-5916  ckuchina@usc.edu  🔗 Chinmay Dattanand Kuchinad  ⚙ chinmaydk99

## EDUCATION

**University of Southern California, Los Angeles**　　　　　　　　　　　　　**Aug 2023 – May 2025**
*Master of Science - Specialisation in Machine Learning and Data Science*　　　　　　　*GPA: 3.61/4.00*

**PES University, Bengaluru**　　　　　　　　　　　　　　　　　　　　　　**Aug 2016 – Jul 2020**
*Bachelors of Technology in Electronics and Communication Engineering*　　　　　　　　*GPA: 8.84/10.00*

## EXPERIENCE

**USC Information Sciences Institute**　　　　　　　　　　　　　　　　　　**Feb 2025 – May 2025**
*Researcher – Edge AI & Distributed Machine Learning*　　　　　　　　　　　*Los Angeles, California*

- Developing **pipeline parallelism strategies for the PipeEdge framework,** creating dynamic partitioning algorithms that **intelligently allocate model layers** based on compute capabilities and network bandwidth across heterogeneous edge devices, **reducing end-to-end latency by 43%** compared to conventional data-parallel approaches
- Implementing both **post-training quantization and magnitude-based pruning methods** with iterative fine-tuning, achieving 60% sparsity in Vision Transformer(Base16) models through **structured channel pruning while preserving critical features for inference tasks**

**Samsung Electronics America**　　　　　　　　　　　　　　　　　　　　**Jun 2024 – Aug 2024**
*Generative AI Application Intern*　　　　　　　　　　　　　　　　*Ridgefield Park, New Jersey*

- Developed a **multi-agent email campaign generation system** using open-source LLMs to orchestrate specialized agents for email marketing campaigns via LangGraph, ensuring consistent brand voice through advanced prompt engineering
- Engineered a **ChromaDB-powered RAG system** for precise product information retrieval, ensuring grounded responses and reducing hallucinations
- Fine-tuned open source LLMs (Llama-3.1, Gemma) on historical marketing campaigns data using PEFT and QLoRA techniques

**Bosch Global Software Technologies Pvt Ltd**　　　　　　　　　　　　　**Jan 2020 – Nov 2022**
*Software Engineer - Machine Learning*　　　　　　　　　　　　　　　　　　*Bengaluru, India*

- Built production ML pipelines (PyTorch) for Honda Japan, analyzing operational data for engine performance optimization; **reduced RAM usage prediction error by 32%**, enhancing predictive maintenance
- Developed an **automated FOTA testing framework,** leveraging Python and SQL to perform in-depth analysis of device telemetry data. Built ML models that **increased flash success rates from 73% to 91%** by identifying key factors influencing success and optimizing for variable network conditions, significantly improving system uptime and reducing update failures
- Formulated and managed **data pipelines orchestrated for the ingestion of ECU(Engine Control Unit) data,** overseeing insights such as EEPROM usage, fault prediction, and memory usage

## Projects

**AI-Powered Multi-Agent System for Sales Automation**　　　　　　　　　**April 2025 – May 2025**
*Orchestrating LLMs using OpenAI agents sdk for End-to-End Sales Email Automation with Guardrails*

- Architected and built an asynchronous, **multi-agent AI system in Python automating end-to-end sales email generation**, integrating LLMs (Gemini, GPT-4o-mini) and SendGrid via OpenAI agentic SDK framework **featuring Pydantic validation and guardrails** (professionalism, brand compliance), designed to scale personalized SDR outreach and enhance communication consistency

**LLM-Based Clinical Diagnosis Prediction**　　　　　　　　　　　　　　　**Aug 2024 – Dec 2024**
*Fine-tuning Llama-3.2 on MIMIC-IV dataset for Next-Visit Disease Forecasting*

- Developed an LLM-powered **sequential disease forecasting model by fine-tuning Llama-3.2** with QLoRA and 4-bit quantization on the MIMIC-IV EHR dataset, achieving **85.43% ROC-AUC and 42.32% PR-AUC** in predicting next-visit diagnoses and demonstrating efficient, high-fidelity clinical predictive modeling

## TECHNICAL SKILLS

**Programming  Scripting:** Python, C++, R, SQL, JavaScript, TypeScript, CUDA
**Machine Learning and Deep Learning:** PyTorch, TensorFlow, Hugging Face, Scikit-learn, XGBoost, LightGBM, MLflow, Grafana, Prometheus
**Data Visualization & Engineering:** Pandas, NumPy, SciPy, Matplotlib, Seaborn, Plotly, Kafka, Spark, Airflow, Snowflake
**Cloud& Backend:** AWS (SageMaker, S3, EKS, Lambda), Docker, Kubernetes, Node.js, Express.js, FastAPI, Flask, Github Actions
**GenAI Tools & Frameworks:** Cursor, Bolt, MCP, Langchain, Langgraph, CrewAI, OpenAI Agents SDK, Google ADK