Rishabh Thapliyal

+1 858-281-1321 | rthapliyal@ucsd.edu | San Diego, CA | Portfolio | LinkedIn

Education

University of California San Diego | *GPA: 3.96/4* — San Diego, CA
Master of Science in Electrical & Computer Engineering | Machine Learning & Data Science — *Sep '24 - Dec '25*
Courses: Generative Models, Scalable Data Systems, Optimizing Deep Neural Networks, Recommender Systems

Indian Institute of Technology (IIT) Bombay | *GPA: 8.4/10* — Mumbai, India
Bachelor of Technology in Chemical Engineering | Minor in Artificial Intelligence & Data Science — *July '18 - May '22*
Courses: Machine Learning, Deep Learning, Machine Learning for Remote Sensing, Mathematical Optimization

Technical Skills

Programming & Tools: Python, C++, SQL, Pandas, NumPy, Dask, Postman, FastAPI, WebSocket
ML & AI Frameworks: PyTorch, TensorFlow, Scikit-learn, Transformers, Large Language Models, NLP, Neural Networks, LLM Fine-tuning, Quantization, Pruning, KV Cache, Reinforcement Learning, FAISS, RLHF
Cloud & Big Data: Ray, PySpark, Hadoop, Vertex AI, Azure AI Studio, Docker, Kubernetes, MLflow, Redis
Agentic frameworks: LangChain, LangGraph, CrewAI, Agno, Autogen, CodeACT, smolagents, LIDA

Work Experience

Qualcomm Inc. — San Diego, CA
Machine Learning/Generative AI intern | *QGenie AI Team* — *Jun '25 - Sep '25*
- Developing, deploying, and scaling **agentic Generative AI** applications such as **Text-to-SQL**, **LogTalk**, & **Slide Generation** to enhance productivity across workflows by serving **50k+** employees.
- **Only intern** to present at first-ever **QGenie AI summit**; deployed the application for Beta testing within **10** weeks.

Walmart Global Tech — Bengaluru, India
Machine Learning Engineer III | *International Global Sourcing & Catalog Team* — *Jun '24 - Sep '24*
- Developed an LLM-powered **Retrieval-Augmented Generative** question answering service for Walmart associates. Collaborated with data engineering, product, and UI/UX teams to ensure seamless accessibility for end users

Machine Learning Engineer II | *Received **Return Offer** after Internship* — *Jun '22 - May '24*
- Engineered **Trend-to-Product** AI tool, leveraging **generative AI** and trend-sensing algorithms (social media, search, and sales data) to reduce the traditional apparel production timeline by up to **18 weeks**.
- Build a Global Item Mapper solution to map product reviews by using product title, descriptions, image features (HSV, SSIM, SIFT, & ORB), and attributes (brand, size, color, etc.) for **200M+** items across US, Mexico, and Canada.
- Created a content quality scoring pipeline to score titles, images, and attributes for **50M+** products in the Walmart's Mexico catalog. Deployed this as an API on **Google Cloud Platform** and scaled it to handle **100+** requests/second

Research Projects

Real-Time Text-to-Speech App | *Audio generation, Real-time streaming, MFA, MathJax* — UC San Diego
Self Learning Project | *Github* — *Aug '25 - Aug '25*
- Designed and implemented a low-latency **bidirectional WebSocket** TTS system with advanced pre/post audio processing using open-source **Kokoro** and **Chatterbox** models, achieving less than **900ms** first text-to-speech latency.
- Enabled robust character-level caption alignment using Montreal Forced Aligner, and dynamic model switching

Unlearning Styles in Diffusion Models | *LoRAs, PyTorch, CLIP, HPS-V2 scores* — UC San Diego
Student Researcher | Advisor: Prof. Nuno Vasconcelos, Statistical & Visual Computing Lab — *Mar '25 - Jun '25*
- Developing **novel** architectural modifications for selective style unlearning in **diffusion** (SDXL-base) models while preserving content fidelity, extending ZipLoRA and UnZipLoRA frameworks.

Wildfire Smoke Detection | *multimodal LLMs, Vision transformers, CNN, LSTM* — UC San Diego
Student Researcher | Advisor: Dr. Mai H. Nguyen, San Diego Supercomputer Center | Github — *Nov '24 - Jun '25*
- Compared **5+** multimodal LLMs (**LLaVA**, **Llama3.2**, **Qwen2.5**, **Gemma3**, **InternVL**) on **500+** fire sequence images and weather data, using prompt engineering and gridding to localize fires and quantify prediction confidence
- Developed scripts for **few-shot learning** and for **fine-tunning** the multimodal LLMs using PEFT techniques (LoRAs)

Achievements

- **Certifications**: GenAI with LLMs, Transformers, Computer Vision, Introduction to Deep Learning, NLP
- **Leadership/Team Work**: IITB Cricket Team **Captain**, Academic **Mentor**, IITB **Placement Team**, **TA** for Physics