

# RISHABH THAPLIYAL

☎ +1 858-281-1321 ✉ [rthapliyal@ucsd.edu](mailto:rthapliyal@ucsd.edu) 🏠 San Diego, CA 📁 [Portfolio](#) 🔗 [LinkedIn](#) 🐙 [Github](#)

## Education

**University of California San Diego** | *GPA: 4.0/4.0* expected graduation Dec '25  
*Master of Science in Machine Learning & Data Science* Sep '24 - Present

**Courses:** Statistical Learning I & II, Data Systems for ML, Advanced Data Mining, Probability & Statistics

**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 & Data Tools:** Python, C++, R, SQL, MATLAB, Pandas, NumPy, Dask, Matplotlib, Seaborn  
**ML & AI Frameworks:** PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, LangChain, CUDA, LoRA, PEFT, Quantization, LLM Fine-tuning, Prompt Engineering, RAGs, Feature Engineering, Model Evaluation  
**Cloud & Big Data:** PySpark, Hadoop, GCP (Vertex AI), Azure AI Studio, Docker, Kubernetes, MLflow, Redis

## Work Experience

**Walmart Global Tech** Bengaluru, India  
*Data Scientist III | International Catalog Team* Jun '24 - Sep '24

- Innovated a novel architecture by integrating **semantic router chains**, multiple LLM sub-chains, & **multi-prompting** techniques in SQL chains library of LangChain to build interactive **Text-to-SQL** queries chatbot with **90%** accuracy
- 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
- Mentored an intern, guiding project execution and skill development while ensuring alignment with team goals

*Data Scientist II | International Global Sourcing & Catalog Team* Jun '22 - May '24

- 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. Scaled this pipeline across multiple **GPUs** (4 T4s) using **Dask** to reduce the model's computation time by **60%**
- Improved the catalog hierarchy for the Walmart's Mexico market by predicting product types from product titles for **10M+** un-navigable items using a max voting **ensemble model** of XGBoost, MPNet, and GPT-3.5
- 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
- Identified trends, seasonality, & cyclicity in the time series data of **20+** KPIs for **8k+** suppliers. Grouped these time series using **Dynamic Time Warping** and created forecasting models for each of them using ARIMAX, Prophet, LightGBM, and **hybrid forecaster model** of Linear Regression and XGBRegressor, achieving **83%** accuracy

*Data Science Intern | International Global Sourcing Team | Awarded a **Return Offer*** Apr '21 - Jun '21

- Pre-processed text data and performed data augmentation to solve for class imbalance using **Snorkel's** transformation functions, leading to **5%** increase in training data and **2%** increase in model performance
- Employed SVM, Multinomial Naive Bayes, **Voting classifier**, and **FastText** to perform multi-class text classification. Used **fuzzy match** scores as an evaluation metric & achieved **1%** performance boost over the company's existing model

## Research Experience & Projects

**Wildfire Smoke Detection** | *multimodal LLMs, LLaVA, Vision transformers, CNN, LSTM* UC San Diego  
*Student Researcher | Dr. Mai H. Nguyen, Associate Director for AI of the WIFIRE Lab at SDSC* Nov '24 - Present

- Working on improving the performance of the multimodal SmokeyNet model for wildfire smoke detection
- Compared **LLaVA** and **Llama** models; and set up LLaVA model on expanse to finetune on fire sequences image data
- Developed scripts to augment the existing dataset with new fire sequences, automating the data pre-processing pipeline

**MeetMinder App** | *Google Vertex AI, Large Language Models, WebVTT* | 🐙 Bengaluru, India  
*GenAI hackthon co-hosted by Google and Walmart* Jul '23 - Aug '23

- Created an application on GCP using **text-bison** LLM to generate meeting's agendas from previous email thread, to live track agendas covered using the zoom transcripts, and to provide minutes of the meet to all meeting attendees

## Achievements

- **Certifications:** [GenAI with LLMs](#), [NLP with Tensorflow](#), [Computer Vision](#), [Introduction to Deep Learning](#), [NLP](#)
- **Leadership/Team Work:** IIT Bombay Cricket Team Captain, Department Academic Mentor, IITB Placement Team