

## Partha Thakuria

Riverside, CA 92507 · +1 (951) 594 -5810 · pthak006@ucr.edu

LinkedIn: <https://www.linkedin.com/in/partha-thakuria> · GitHub: <https://github.com/parthathakuria>

## Academic Highlights

### The University of California, Riverside, CA

2022 – 2024

Master of Science in Computer Engineering (3.81 GPA)

*Focus:* Machine Learning, Scientific Computing

*Projects:* Machine Learning, Data Mining, NLP

### The University of California, Riverside, CA

2024 – Present

PhD in Computer Engineering

*Focus:* Machine Learning, Information Theory

## Technical Proficiencies

**Programming Languages & Framework Tools:** C | C++ | Python | SQL | Bash Scripting | Shell Scripting | Git | JIRA | TensorFlow | PyTorch | Keras | numpy | pandas | matplotlib | sklearn | Huggingface API | OpenAI API | SageMaker | S3 | Lambda | AWS | Docker | CUDA | Django | Flask | REST API | Fast API | TFX

**Statistical Tools:** Regression Analysis | Time Series Analysis | Bayesian Analysis | Markov Decision Process | Hypothesis Testing | Generative & Discriminative Models

**Machine Learning:** Supervised Learning | Unsupervised Learning | Deep Learning (NLP, Computer Vision)

## Professional Experience

### System Engineer, Unisys India, Bangalore, India

2018 – 2019

Implemented software for automating deployment of micro-segmentation security systems. Coordinated with team members for project development, bug resolution, and UI/API enhancements.

## Teaching Experience

### Teaching Assistant, CS010A - Introduction to Programming (C++), UC Riverside

Duties included assisting students with programming assignments, grading examinations, and proctoring. Provided guidance in foundational C++ concepts and problem-solving strategies.

### Teaching Assistant, CS170 - Introduction to Artificial Intelligence, UC Riverside

Conducted discussion sessions on AI principles, held office hours for assignment and project assistance, and contributed to the design of course projects and assignments. Managed and configured the autograder for Python project submissions.

## Key Projects

### Data Mining, AI, and NLP Course Projects, UC Riverside, CA

**Data Mining Project:** Developed a neural network-based solution for identifying Higgs boson signals. Employed regularization and ensemble approaches.

**NLP Project:** Conducted sentiment analysis using models like BERT-UDA, XLNet, and RoBERTa large. Achieved better accuracy in GPU settings than benchmark implementation.

**Spatial Computing Project:** Developed a web-based interface for safe path mapping in NYC. Utilized ML algorithms and CUDA optimization.

### Machine Learning Course Projects:

(a) *GPT-3 Calculator:* Explored the GPT-3 pre-trained model's capabilities in numeric tasks without fine-tuning. Experimented with various prompting techniques and statistical analysis to understand model behavior and improve accuracy. Faced challenges with skewed data, addressed by employing advanced sampling techniques for predictive modeling.

(b) *Semantic Segmentation with ResNet*: Trained a ResNet model on the VOC 2007 dataset for semantic segmentation. Analyzed model confidence using calibration plots and enhanced performance with L2 regularization, Monte Carlo simulation, and data augmentation techniques.

## Research Projects

**Climate Model Emulation**, UC Riverside

*Supervisor*: Professor Andy Ridgeway, Earth and Planetary Science Department

Engaged in the implementation of machine learning models, including Gaussian Processes and CNN-LSTM, for emulating climate models. Work involved standardizing the ClimateBench repository and enhancing climate model predictions through data-driven approaches.

**Echo Diffusion: Analyzing Diffusion Models with Information Theory**, Master's Project under Professor Gregory Ver Steeg

*Related Work*: "Echo Diffusion: Analyzing Diffusion Models with Information Theory"

This project explores the integration of Echo Noise into the diffusion model framework to potentially enhance the performance of generative models. Functions were implemented to calculate Echo Noise in a data-driven fashion and mutual information or echo loss calculations. The results showed that the cold diffusion model with Echo Noise achieved superior performance in terms of reconstruction on the MNIST dataset, validating the effectiveness of Echo Noise in capturing meaningful latent representations. Further investigation and optimization are needed for more complex datasets like OMNIGLOT. For more details, refer to the project document.

## Education

**Bachelor of Technology in Computer Science and Engineering**

National Institute of Technology Silchar, Assam, India

## Certifications

Machine Learning Specialization, Coursera

Deep Learning Specialization, Coursera