Aayush Gupta

Houston, Texas • 505-313-2769 • ayushguptastudent@gmail.com • LinkedIn • GitHub • Google Scholar

PROFESSIONAL EXPERIENCE

University of Houston, Houston · *May 2024 - Present*
Research Assistant, Advisor: Gopal Pandurangan

Pacific Northwestern National Lab, Richland, Washington · *May '23 - Aug '23*
Ph.D. Intern: Data Scientist, Advisor: Mahantesh Halappanavar

Center for Advanced Research Computing, Albuquerque, New Mexico · *Aug '21 - Dec '21*
Graduate Research Assistant, Advisor: Matthew Fricke

Accessibility Resource Center, Albuquerque, New Mexico · *Aug '21 - Dec '21*
Program Support Staff, Coordinator: Frankie McQuerry

Timbl (RI Networks Pvt. Ltd.), Gurugram, Haryana, India · *May'18 - Sept'18*
Intern: Network Analyst, Advisor: Kapil Dev Kumar

EDUCATION

University of Houston, Houston, TX, USA · *2022 - Present*
Ph.D. in Computer Science, Advisor: Gopal Pandurangan

University of New Mexico, Albuquerque, NM, USA · *2019 - 2021*
M.S. in Computer Science, Advisor: Thomas Hayes

College of Engineering Roorkee (UTU), Roorkee, India · *2014 - 2018*
B.Tech. in Electronics and Telecommunication

PUBLICATIONS

Academic Network: Erdős Number **3**

Conference/Journal Papers

- *Fully-Distributed Construction of Byzantine-Resilient Dynamic Peer-to-Peer Networks* by **A Gupta** and G Pandurangan. *arXiv*

- *Brief Announcement: Fully-Distributed Construction of Byzantine-Resilient Dynamic Peer-to-Peer Networks* by **A Gupta** and G Pandurangan. *Brief Announcement, In the Proceedings of the 2025 ACM Symposium on Parallelism in Algorithms and Architectures (SPAA '25)*

- *Energy-Efficient Maximal Independent Sets in Radio Networks* by D Banasik, V Dani, F Dufoulon, **A Gupta**, TP Hayes, and G Pandurangan. *In 39th International Symposium on Distributed Computing (DISC 2025). Leibniz International Proceedings in Informatics (LIPIcs), Volume 356, pp. 14:1-14:24, Schloss Dagstuhl – Leibniz-Zentrum für Informatik (2025).*

- *Brief Announcement: Energy-Efficient Maximal Independent Sets in Radio Networks* by D Banasik, V Dani, F Dufoulon, **A Gupta**, TP Hayes, and G Pandurangan. *Brief Announcement, In Proceedings of the 2025 ACM Symposium on Principles of Distributed Computing (PODC '25).*

- *Wake up and join me! An energy-efficient algorithm for maximal matching in radio networks* by V Dani, **A Gupta**, TP Hayes and S Pettie. *35th Int. Symposium on Distributed Computing, (DISC '23).*

- *Brief Announcement: Wake Up and Join Me! An Energy Efficient Algorithm for Maximal Matching in Radio Networks* by V Dani, **A Gupta**, TP Hayes, and S Pettie. *Proceedings of the 40th ACM Symposium on Principles of Distributed Computing (PODC '21).*

POSTERS & TALKS

- *Software Vulnerability Detection: A Grand Challenge for AI* by **A Gupta**, R Verma, and A Dunbar. *46th IEEE Symposium on Security and Privacy (IEEE S&P '25), Poster Session.*

- *Fully-Distributed Construction of Byzantine-Resilient Dynamic Peer-to-Peer Networks* by **A. Gupta**, G. Pandurangan. *36th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA '25), Portland, OR. Presented talk in full paper session.*

- *Fully-Distributed Construction of Byzantine-Resilient Dynamic Peer-to-Peer Networks* by **A. Gupta**. Poster presented at the **2025 Computer Science PhD Research Showcase**, University of Houston, Nov 5, 2025.

- *Comparative Analysis of Approaches to Software Vulnerability Detection on High-Quality Datasets* Poster presented at **U.S. Department of Transportation CYBER-CARE OST-R Visit**, University of Houston, Oct 2024. Showcased performance of deep learning models (DistilBERT, TextRCNN) on MVDSC and NVD datasets for software vulnerability detection.

UNDER SUBMISSION/PREPARATION

- *A Comprehensive Taxonomy for Code Datasets with Applications to Software Vulnerability Detection* by **A Gupta**, AP Dunbar, R Verma.

- *A Strongly-Polynomial Time Algorithm for Learning Noisy Linear Threshold Functions: Simpler, Faster, and Parallel* by R Das, **A Gupta**, C Neibert, G Panurangan, R Verma.

SELECTED PROJECTS

Spontaneous Facial Expression Recognition
- Developed a machine learning model to recognize facial expressions in images by extracting Histogram of Oriented Gradients (HOG) and Local Gabor Features (LGF).
- Utilized classical classifiers and neural networks like CNNs and RNNs, to identify different facial expressions.
- Applied ensemble methods like AdaBoost to improve classical model performance and evaluated the models.
- Fine-tuned the neural networks by adding custom layers, LoRA technique, and tuning hyperparameters.

Machine Learning Engineer - Natural Language Processing Project
- Developed a question-answering system using the CommonsenseQA dataset and fine-tuned RoBERTa model, enhancing understanding of contextual nuances in natural languages.
- Engineered data preprocessing pipelines and optimized hyperparameters to improve model performance, achieving competitive accuracy and top-k accuracy metrics.
- Utilized machine learning techniques including transfer learning and performance tuning to refine model predictions and evaluate them using metrics like ROC-AUC.
- Leveraged Python, PyTorch, and Hugging Face's Transformer library to implement, train, and evaluate the model, gaining hands-on experience in cutting-edge NLP technology.

Deep Learning Project: Fashion MNIST Image Classification

- Developed a neural network model to classify fashion items using the Fashion MNIST dataset.
- Used Python, TensorFlow, Keras, efficientnetv2-xl-21k model.
- Experimented with various hyperparameters to optimize model performance.

Bots: A boon or a bane
- Simulated a DDoS attack using BoNeSi simulator.
- Produced ICMP, UDP, and TCP packets to flood a network.
- Total of 1150610 packets were sent in a period of 10 seconds.
- Evaluated the model's performance using metrics such as Top-k accuracy, classification report, and confusion matrix.

Improving Parallel Processing using Deep Learning
- Used Torchbraid which is a Python package designed for interaction with PyTorch library.
- Trained and analyzed neural networks in a layer-parallel fashion

Code Cloning Detector
- Code clone detection program that uses a machine learning classifier trained on the BigCloneBench dataset.
- Proved robust as well as asymptotically energy efficient in complete and arbitrary network topology.

Conceptual Design document for Computer Science Student Recruitment and Retention
- Designed an EER model and its schema for the Computer Science department at the University of New Mexico.
- This model considered the information and the relationships about the students, teachers, and different programs ongoing in the department, courses, advisors, assistantships, tutors, research papers, internships, surveys, awards, mentors, and other related special activities in the department using MySQL.

Implementation of Simply-Typed Lambda Calculus
- Implemented the complete Simply-Typed Lambda Calculus in Haskell including α-conversion, β-conversion, and type evaluations.

TECHNICAL SKILLS

Programming Languages: Python, C, C++, MATLAB, Haskell, Bash, SQL, LaTeX

Machine Learning / Deep Learning: PyTorch, TensorFlow, HuggingFace Transformers, Scikit-learn, XGBoost, LightGBM, Optuna, LoRA, PEFT, RAG, RLHF, CNNs, RNNs, LSTMs, BiLSTMs

Large Language Models / NLP: LLMs (e.g., GraphCodeBERT, GPT-Neo, T5, Mixtral, CodeBERT, Falcon, Llama), Transformers, vLLM, LangChain, LlamaIndex, BERT-based fine-tuning, Few-shot and Zero-shot Prompting

Distributed Systems / HPC: MPI, OpenMP, Slurm, CUDA, PBS, Kubernetes, Docker, Ansible, PBS, Emacs, Vim, Git

Cloud / Deployment: AWS EC2/S3, Google Cloud Platform (GCP), Azure, HPC clusters (local and university-level), cPanel Web Hosting

Data / Feature Engineering: Pandas, NumPy, NetworkX, OpenCV, AST (Abstract Syntax Tree) Processing, CFG Graph Analysis, Cleanlab, TF-IDF, Cosine Similarity, Deduplication

Security & Vulnerability Detection: Static Code Analysis, AST-aware masking, Code Similarity Matching, Label Noise Detection (Cleanlab), Dataset Normalization, Cross-dataset Contrastive Learning

Databases / Web / Visualization: PostgreSQL, MongoDB, SQLite, HTML/CSS, JavaScript, Matplotlib, Seaborn, Plotly, Tableau, Matplotlib, Seaborn, Excel

TEACHING EXPERIENCE

- *Teaching Assistant*, Algorithms and Data Structures, UH Computer Science (Spring 2025, 2024, 2023)

- *Teaching Assistant*, Database Systems, UH Computer Science (Fall 2024)

- *Teaching Assistant*, Introduction to Automata and Computation, UH Computer Science (Fall 2023)

- *Teaching Assistant*, Data Science , UH Computer Science (Spring and Summer 2022)

- *Teaching Assistant*, Operating Systems, UNM Computer Science (Spring 2021)

- *Teaching Assistant*, Intermediate Programming, UNM Computer Science (Spring 2020)

- *Grader*, Dynamics, Department of Mechanical Engineering, UNM (Fall 2020)

COURSEWORK

- Algorithms and Data Structures

- Randomized Algorithms

- Machine Learning

- Computer Networks

- Introductions to Parallel Processing

- Database Management Systems

- Introduction to Deep Learning *(Hewlett Packard Enterprise Data Science Institute)*

- Artificial Intelligence

- Introduction to Cybersecurity

- Program Analysis and Testing

- Research Methods in Computer Science

- Complex Adaptive Systems

- Software Foundations

- Operating Systems

- Geometric and Probabilistic Methods in Computer Science

- Introduction to Theory of Computation

- Advanced Numerical Analysis

- Introduction to Machine Learning *Hewlett Packard Enterprise Data Science Institute*

- Experimental Methods in Computer Science