

M S RAAGHAVAN

Hyderabad, India | +91 6303800138 | ms.raaghavan@gmail.com
[linkedin.com/in/raaghavan-ms-2b1245298](https://www.linkedin.com/in/raaghavan-ms-2b1245298) | github.com/msraaghavan | leetcode.com/u/MSR_2005 | [Portfolio](#) | [Certifications](#)

SUMMARY

Computer Science and Engineering undergraduate (B.Tech) and machine learning researcher with three peer reviewed publications (IEEE, Springer, Elsevier) in graph neural networks, multimodal fusion and natural language processing (NLP). Designed, trained and rigorously evaluated deep learning systems in PyTorch and PyTorch Geometric, from diffusion models and generative adversarial networks (GANs) to large language model (LLM) fusion, and engineered distributed data pipelines in Apache Beam.

EDUCATION

Hyderabad Public School | Class 10 (ICSE) | Hyderabad | Jul 2021
Sri Chaitanya Junior College | Class 12 (MPC) | Hyderabad | May 2023
Amrita Vishwa Vidyapeetham | Bachelor of Technology, Computer Science and Engineering | Kerala | Aug 2023 to Present
Relevant Coursework: Machine Learning, Artificial Intelligence, Natural Language Processing, Graph Mining, Data Structures and Algorithms, DBMS, Operating Systems

TECHNICAL SKILLS

Languages: Python, Java, C, JavaScript, SQL, Bash
Deep Learning: PyTorch, PyTorch Geometric, TensorFlow, Hugging Face Transformers, Scikit Learn, NumPy, Pandas, ONNX Runtime
Modeling: Graph Neural Networks (GAT, SE(3) equivariant), Transformers, GANs, Diffusion Models, Computer Vision, cross attention fusion, contrastive learning (InfoNCE), LoRA, quantization
Data and Distributed Systems: Apache Beam, batch and streaming pipelines, spatial partitioning, Parquet, pyarrow, REST API design
LLM and NLP: LLaMA, BioGPT, BioBERT, LangChain, RAG, FAISS, ChromaDB, SentenceTransformers, OpenAI API
Tools and Infrastructure: Git, Docker, Linux, GCP, Flask, Streamlit, multi GPU training (A100), Weights and Biases, MySQL, MongoDB

EXPERIENCE

Dev Dungeon | Data Analyst Intern | Kerala | Aug 2024 to Jan 2025

- Built interactive Streamlit dashboards that consolidated key business metrics into a single view for stakeholders, replacing manual spreadsheet reporting with a self service view of performance.
- Analyzed customer datasets in Python and Pandas to surface usage and retention trends, delivering insights that informed product prioritization and improved the efficiency of data driven decision making.
- Contributed to the startup's AI and data science initiatives, supporting its product direction and strategic development.

PROJECTS

GeoBeam: Distributed Spatial Query Engine | Apache Beam, Python, Flask, Leaflet | [Code](#) | [Live Demo](#)

- Built a distributed spatial query engine on Apache Beam implementing four canonical spatial operations (range queries, spatial joins, k nearest neighbor search and partition counts) as composable PTransforms.
- Implemented three pluggable partitioning strategies (uniform grid, depth 4 quadtree, and reservoir sampled adaptive grid), evaluated over 12.7M NYC TLC taxi trip records.
- Recovered point geometry from source data containing only zone IDs via TLC shapefile joins and EPSG:2263 to WGS 84 reprojection; served results through a Flask REST API and a deployed Leaflet interface.

HumanDiT: Controllable Face Generation | PyTorch, Diffusion Transformers, GANs, ONNX, WebGPU | [Code](#) | [Live Demo](#)

- Trained a 183M parameter diffusion transformer with flow matching in a frozen DINOv3 latent space (RAEv2) on FFHQ, generating 256x256 faces steerable by 12 attributes; reached an FID of 3.78 on 50k samples and 89.9% attribute control accuracy.
- Distilled it into a single step GAN student (FID 20.0 vs 23.0 for a flow distillation baseline) after diagnosing mode collapse and a discriminator noise schedule bug; audited memorization against all 70k training faces.
- Deployed in the browser with ONNX Runtime Web and WebGPU (4 and 8 bit quantized): 0.54 s per face on a laptop GPU.

PUBLICATIONS

MatFusionNet: Transformer-GNN Fusion for Zero-Shot Band Gap Prediction | [Paper](#)
ICMLDE 2025, *Procedia Computer Science* (Elsevier) | 2nd author of 5

- Jointly developed a multimodal architecture fusing a 6 layer message passing GNN over periodic crystal graphs with a Transformer text encoder via gated cross attention.
- Achieved 0.271 eV MAE on the Materials Project benchmark (84,312 compounds), a 7.2% reduction over MEGNet, statistically significant at $p < 0.01$ across five seeds.

GLaMP: Graph-Latent Alignment for Malicious Phishing Detection | [Paper](#)
IEEE ICCCNT 2025, IIT Indore | 3rd author of 5

- Built a Graph Attention Network over a heterogeneous metadata graph (WHOIS, TLS certificates, redirect chains, threat feeds) with learnable exponential decay edge weighting, aligned to a LoRA fine tuned LLaMA 7B encoder via an InfoNCE contrastive objective.
- Reached 91.3% accuracy and 0.90 AUC, 4.6 points above a GCN+BERT late fusion baseline.

SLPIM: Structural-Linguistic Protein Interaction Modelling | [Paper](#)
CONFLUENCE 2025, *Springer LNCS* | 3rd author of 5

- Integrated SE(3) equivariant GNNs over residue proximity graphs with fine tuned LLaMA 2 and BioBERT/BioGPT encoders across PDB, STRING, UniProtKB and BioGRID, reaching 90.36% annotation accuracy (BLEU 4: 0.56, ROUGE L: 0.48) and a 13% gain over single modality baselines.

LEADERSHIP & OUTREACH

- Lakshya, Student Social Responsibility Program (2025): jointly delivered a three hour career guidance and AI ethics workshop to Class 9 and 10 students in Oachira, Kerala. | [Report](#)