Syed Affan Daimi

+91 939 045 6483 | saffand03@gmail.com | LinkedIn | GitHub | Portfolio | Hyderabad, India

Education

International Institute of Information Technology, Hyderabad — Hyderabad, India
M.S. in Computer Science; ***GPA: 9.0/10.0*** — *Aug 2025 – Present*
Electives: Language Models and Agents, Applications of LMs, Data Systems, NLP Evaluation Methods

Projects: TTGemma – Tiny Trilingual Gemma, Graph DB from Scratch, When Does the Judge Know Best? A Multi-Axis Study of LLM-as-Judge Reliability for News Summarisation

Indian Institute of Technology, Madras — Chennai, India
B.S. in Data Science and Applications; ***GPA: 8.0/10.0*** — *Graduated Aug 2025*
Electives: Algorithms for Data Science, AI: Search Methods, Design Thinking, Market Research

Muffakham Jah College of Engineering and Technology — Hyderabad, India
Bachelor of Engineering; ***CGPA: 8.0/10.0*** — *Graduated Jun 2024*

Experience

International Institute of Information Technology — Hyderabad, India
Associate Researcher — *Jun 2024 – Present*

- Engineered an automated genomics report system for startup **GenePoweRx**, building a data ingestion pipeline to merge unstructured VCF genomic files with structured physician logs, replacing a manual process and reducing reporting turnaround by 60%.
- Developed a scalable visual analytics system to interpret community evolution in large, time-varying networks via a novel algorithmic approach for classifying and visualizing temporal dynamics. Validated efficacy through a formal User Study.
- **Teaching Assistant**: Data Visualization course; facilitating 50 students on D3.js, visualization design and grammar.

Razor Group GmbH — Berlin, Germany
Data Science Intern — *Nov 2023 – May 2024*

- Improved demand forecasting accuracy by 20% by implementing time-series clustering to segment products by sales velocity.
- Deployed a serverless AWS Lambda pipeline to track 500,000+ competitor products in real-time, feeding live insights into BI dashboards.
- Enhanced model transparency via interpretable visualizations, enabling stakeholders to audit feature relevance and model sensitivity.

Indian Institute of Technology, Hyderabad — Hyderabad, India
Research Intern — *May 2023 – Jul 2023*

- Built an object detection system for autonomous driving on 19,800+ radar frames across 6 object categories in urban environments.
- Elevated F1-score from 60% to 71% via a custom Vision Transformer (ViT) in PyTorch, outperforming 7 conventional CV backbones.

fellowship.ai — Remote
Data Science Intern — *Oct 2022 – Jan 2023*

- Optimized an image similarity search pipeline by over 95%, cutting query time from 5+ minutes to under 45 seconds across 5,000+ items using CLIP embeddings, batch processing, pre-computed storage, and PaCMAP dimensionality reduction.

Selected Publications

- [**Under Review**] **Syed Affan Daimi**, Gaurav Sarkar, Kamalakar Karlapalem. "Glass Doors: Competency as a Behavioral Dimension for Interpreting Agent Failure." Submitted to NeurIPS 2026 Workshop IAB.
- [**Under Review**] Gaurav Sarkar, **Syed Affan Daimi**. "Mask-Pattern Position Augmentation: Training-Free Long-Context Extrapolation for Diffusion LLMs." Submitted to NeurIPS 2026 Workshop DiffuLM.
- [**Preprint**] Gaurav Sarkar, **Syed Affan Daimi**, Jay Gala, Subarna Tripathi. (2026). "SG-Blend: Learning an Interpolation Between Improved Swish and GELU for Robust Neural Representations." *arXiv preprint arXiv:2505.23942*.

- **Syed Daimi***, Garima Jindal, Kamal Karlapalem. "DyNetSpin: Analysis of Dynamic Networks via Spiral Layouts." EuroVis 2026 (Posters and Demos).
- **[Preprint] Daimi, S. A.*** (2024). A Scalable Data-Driven Framework for Systematic Analysis of SEC 10-K Filings Using Large Language Models. *arXiv preprint arXiv:2409.17581.*
- **Daimi, S. A.***, Askari, A. A.*, et al. (2025). Leveraging Machine Learning Models For Proactive Disaster Forecasting. In *Proceedings of ICAET 2025.*
- Iqbal, A.*, **Daimi, S. A.**, & Chari, K. M. (2023). Performance Efficient and Fault Tolerant Approximate Adder. *Journal of Electronic Testing, 39*(5), 571–582.

Projects

VidhyaDojo: Multi-Agent Teacher Training Simulator
- Built a multi-agent classroom simulator for novice teachers to practice by addressing LLM's tendency to answer from expert knowledge rather than behave as genuine learners. Modeled each LLM student's evolving knowledge with a mastery-scored graph, developed contrastive log-probability probes to detect gaps and misconceptions, and designed a mastery-weighted logit processor to constrain generated questions at the token level using that same knowledge state. Achieved genuine-gap-detection AUC of 0.814–0.924 across three domains and 14% higher inter-student personalization in ablation.

Pharmaceutical Supply Chain Optimization | *Project Report*
- Digitized manual records into Python, building the business's first structured dataset for inventory and credit tracking. Engineered a KNN Regressor (R^2=0.72) to forecast vendor retrieval times, driving a data-backed policy to discard faulty stock exceeding 90 days and optimize storage costs. Applied Chi-Square testing ($p < 0.01$) to customer credit logs, surfacing a significant correlation between borrower type and default rates. Proposed a B2B-only credit pivot projected to recover a 19.5% revenue loss attributed to high-risk individual defaulters.

TTGemma: 148M Parameter Trilingual Language Model | *Repository*
- Architected a decoder-only transformer from scratch in PyTorch with modern LLM primitives (RMSNorm, RoPE, SwiGLU). Engineered a 4-GPU Distributed Data Parallel training pipeline, resolving memory and I/O bottlenecks via a custom memory-mapped data cycler to stream a 3B-token corpus. Applied LoRA PEFT to adapt the pretrained model for downstream multilingual classification.

Verdure: Plant Health Diagnostics
- Developed a PyTorch computer vision model to detect plant disease, leveraging a dataset expanded to 85k images via augmentation. Engineered a Streamlit interface for user interaction, securing 1st place at Envisage 2023 and 3rd place at Innovatia Panoply 2023 against 200+ participants.

Services

Stanford University - Code In Place Remote
Section Leader & Research Contributor *2025*
- Mentored a global cohort of 20 students through Stanford's CS106A curriculum, leading weekly virtual sessions and providing personalized feedback on Python fundamentals.
- Contributed to educational AI research by evaluating and annotating model-generated feedback, supporting a model featured in a SIGCSE 2026 publication for achieving performance parity with GPT-4.

Open Source Contributions
- **University HPC Tooling:** Built a custom CLI suite automating SLURM node allocation, SSH tunneling, and remote IDE initialization into a single command, eliminating HPC workflow friction. Implemented state-management and teardown scripting to cleanly cancel distributed jobs, kill port forwards, and sync artifacts to permanent storage.
- **MAPIE (scikit-learn-contrib):** Fixed group-based cross-validation compatibility in `_MapieRegressor` for conformal prediction-based uncertainty quantification. Added targeted unit tests and improved documentation for reproducibility.
- **Vizro (McKinsey):** Enhanced the open-source visualization library by contributing targeted website improvements and authoring a new user documentation tutorial.
- **Hugging Face x scikit-learn:** Modernized official scikit-learn tutorials by deploying interactive Gradio versions on Hugging Face Spaces during a contributor sprint, iterated on community feedback.

Technical Writing & Outreach
- Authored technical tutorials and guided projects on Medium, cultivating a consistent readership of 500+ monthly viewers by making complex ML frameworks more accessible.