Abhay Kumara Sri Krishna Nandiraju

abhay2001sri@gmail.com | abhaynandiraju | srikrish2812

Education

- **University of Pittsburgh** *Aug 2025 - Present*
  Doctor of Philosophy in Intelligent Systems Pittsburgh, USA
- **University of Arizona** *Aug 2023 - May 2025*
  M.S in Information Science: Machine Learning (GPA: 4.00/4.00) Tucson, USA
- **Indian Institute of Technology (IIT), Tirupati** *Aug 2019 - July 2023*
  B.Tech in Electrical Engineering (GPA: 8.84/10.00 = 3.80/4.00) Tirupati, India

Publications

C=Conference, J=Journal, P=Patent, S=In Submission, T=Thesis

[C.1] A.K.S.K Nandiraju, et al. (2025). **Automated Feedback Loops to Protect Text Simplification with Generative AI from Information Loss**. *Intelligent Systems Conference 2025*.

[T.1] Abhay Nandiraju (2023). **Efficient Data Augmentation for Tiny Drone Detection System**. Thesis submitted towards my Bachelor's final year project.

Research and Publication Experience

- **Intelligent Computing for Clinical Imaging (ICCI) Lab, University of Pittsburgh** *Aug 2025 - Present*
  PhD Graduate Student Researcher Pittsburgh, PA
  - I research on multimodal deep learning architectures, fusion techniques and representation learning to capture interactions and shared information between different modalities.
  - Currently, I model temporal patterns in Breast DCE-MRI, spatial patterns in Mammograms, and inter-modal interactions to predict breast cancer recurrence.
- **Department of Management Information Systems, University of Arizona** *Sep 2024 - May 2025*
  Graduate Research Assistant - NLP Researcher Tucson, Arizona
  - Published an automated feedback loop framework to mitigate information loss in Generative AI-based biomedical text simplification.
  - Engineered a system to identify and re-insert missing biomedical named entities, increasing content overlap by 29.5% and semantic alignment by 8.2%.
  - Demonstrated that re-inserting all missing entities yielded the highest semantic alignment (0.9162 cosine similarity) and content overlap (0.6555 ROUGE-1) at the document level.
- **Visual Information and Signal Analysis Lab (ViSAL)** *Sep 2022 - Aug 2023*
  Computer Vision Researcher IIT Tirupati, India
  - Authored B.Tech thesis on a data augmentation technique for tiny object detection, addressing critical data scarcity for drone surveillance under the guidance of Dr. Rama Krishna Gorthi.
  - Designed an image composition pipeline to generate synthetic drone samples, using SiamMask to create binary segmentation masks for copy-paste augmentation.
  - Trained YOLOv5 and YOLOv8 models on the synthetic dataset, achieving 0.992 mAP outperforming models trained on the original ICPR dataset (0.984 mAP).

Applied AI and Engineering Experience

- **WML IT Solutions Pvt Ltd** *Aug 2023 - Apr 2024*
  Associate Consultant - AI Engineer Remote
  - Engineered a multi-source Retrieval-Augmented Generation (RAG) system enabling natural language querying across disparate data silos.
  - Developed a natural language-to-SQL agent and integrated Langchain with GPT-3.5 and SQLCoder to interact with Postgres, non-relational DBs, and PDFs.
  - Utilized Postgres as a vector database for efficient semantic search and retrieval of text document embeddings.
- **IIT Tirupati Navavishkar I-Hub Foundation** *Oct 2022 - Aug 2023*
  Chanakya Fellow, Computer Vision Engineer IIT Tirupati, India
  - Secured a 100,000 INR grant to architect, build, and deploy a portable real-time tiny drone detection system.
  - Engineered and optimized a lightweight YOLOv8-nano model, integrating mask-based data augmentation method from my B.Tech thesis.
  - Successfully deployed the end-to-end detection pipeline on a Raspberry-Pi 4, achieving low-inference times for real-world portability.
- **Visual Information and Signal Analysis Lab(ViSAL)** *May 2022 - Sep 2022*
  Computer Vision Research Intern IIT Tirupati, India
  - Investigated super-resolution techniques (BSRGAN, Bicubic interpolation) to enhance small object detection ($< 32 \times 32$ px) in Faster-RCNN and YOLOv5 models.
  - Integrated super-resolution into training pipeline, achieving an 8% mAP improvement for the small objects in COCO dataset ($> 200,000$ images)

SELECTED PROJECTS

- **ReasonLLM: Enhancing Mathematical Reasoning with GRPO** *May 2025*
  PyTorch, Transformers, Unsloth
  - Implemented a training pipeline using Group Relative Policy Optimization (GRPO) via the Unsloth framework on the GSM8K dataset to enhance the mathematical reasoning capabilities of the base 1B-parameter Gemma-3 model.
  - Designed composite reward functions for evaluating output format, step-by-step logic, and exact answer extraction to accurately guide policy updates and penalize inconsistencies.
  - Demonstrated a 5% improvement in zero-shot accuracy on complex reasoning tasks and reduced response errors by 10%, validating reinforcement learning as an effective alignment strategy for small language models.
- **Biomedical Knowledge Agent - PubMed Agentic RAG** *April 2025*
  LangChain, LangGraph, Weaviate, OpenAI
  - Architected an Agentic RAG system over PubMed Central using LangGraph, dynamically routing queries between domain-specific vector search and external Wikipedia retrieval via iterative ReAct prompting.
  - Leveraged PubMedBERT-MS-MARCO for dense vector embeddings, providing a local Weaviate database with tuned HNSW indices and cosine similarity for quick retrieval and similarity search.
  - Engineered a token-aware semantic chunking and XML text-extraction pipeline, explicitly parsing dense PubMed hierarchies to extract metadata while preserving LLM context windows.
- **AI Hedge Fund - Hack Arizona 3rd place** *Mar 2025*
  Python, Fast API, Amazon Bedrock, Claude
  - Architected a multi-agent trading simulator using LangGraph, orchestrating 8+ specialized LLM agents (Valuation, Sentiment, Technicals) to concurrently process financial data and generate deterministic portfolio allocations.
  - Engineered a vectorized backtesting engine in Pandas/NumPy to evaluate long/short equity strategies, rigorously accounting for margin and cost-basis while computing financial metrics like Sharpe ratios and max drawdowns.
  - Implemented algorithmic quantitative valuation models (Discounted Cash Flow, Buffett's Owner Earnings) that dynamically estimate intrinsic value from TTM financial data to enforce risk limits and optimize position sizing.

SKILLS

- **Programming Languages:** Python, Bash, LaTex
- **Machine Learning and Data Science:** PyTorch, Scikit-Learn, Transformers, NumPy, Pandas, Langchain
- **Database & Tools:** HPC, Git, GitHub, Hugging Face, Docker, Postgres, MySQL, Postman, Vim

HONORS AND ACHIEVEMENTS

- **Distinguished Graduate Scholar, College of Information Science, University of Arizona** *May 2025*
- **Information Science Scholarship ($5,000), University of Arizona** *Aug 2024*
- **Joint Entrance Examination (JEE) Advanced, All India Rank 3574 (Top 0.015%)** *May 2019*
- **Joint Entrance Examination (JEE) Mains, All India Rank 3722 (Top 0.003%)** *May 2019*