

SAIDINESH POLA

Ecity, Bangalore

📞 9542491303 ✉️ saidines12@gmail.com 🔗 [linkedin.com/in/saidinesh-pola-275262130](https://www.linkedin.com/in/saidinesh-pola-275262130) 🌐 saidineshpola.github.io/

Experience

Infosys

June 2021 – Current

Specialist Programmer L3 (AI/ML)

Bangalore, India

- **Gen-AI Recruitment Platform:** Led the development of Gen-AI recruitment platform's APIs, implementing NLP techniques including Named Entity Recognition for resume parsing, Evaluation Form generation, candidate recommendation algorithms, reducing effort and time.
- **Multimodal Image Retrieval Business Assistant:** Implemented advanced computer vision and RAG answering machine, including multimodal image retrieval systems and chatbot functionality with noval synthetic dataset generation (RAFT) for llama3 finetuning to improve business intelligence.
- **Automated ML Pipelines:** Developed automated ML pipelines for candidate/job recommendations, Jira proxy agent and evaluation form generation using RAG with fine-tuned Llama3/GPT4 models.
- **MLOps Practices:** Utilizing WANDB for model tracking, hyperparameter tuning, and experiment management, ensuring reproducibility and performance consistency, and follow-up with the Responsible AI office.

AICrowd

Jan 2021 – May 2021

Research Fellow

Remote

- **WHO Schistosomiasis Prevention:** Created SOTA CV model for WHO's schistosomiasis prevention, achieving 52.5 Macro F1 score, and built an accompanying web app using FLASK, demonstrating ability to deliver end-to-end ML data collection system.

IIT GUWAHATI

April 2020– Jan 2021

Researcher/ Assistant

Guwahati

- Focusing on signal processing and communication systems, my responsibilities included developing and simulating communication algorithms using MATLAB, conducting experiments, and analyzing data to optimize system performance.

Education

Indian Institute of Technology, IIT Guwahati

June 2019 – May 2021

Communication Engineering, MTech

GPA 7.47

National Institute of Technology, NIT Raipur

June 2014 – May 2018

Electronics and Communication Engineering, BTech

Technical Skills

Programming: Python, TypeScript, PostgreSQL, C++, CUDA

ML/AI Frameworks: Scikit-learn, PyTorch, TensorFlow, Transformers, CLIP, RAG, LLaMAIndex

ML Engineering: Model optimization, MLOps, Docker, Kubernetes, Production deployment

Cloud & Tools: AWS, GCP, Git, Gitlab, WANDB, Colab

Specializations: Machine Learning, NLP, Computer Vision, Optimizations, Probability, Reinforcement Learning, Code Generation Models

Certifications: CUDA using Python (Nvidia), Machine Learning, Deep Learning (Coursera), Reinforcement Learning (Stanford), Gen-AI (Infosys)

Achievements

Hackathons & Competitions

Spring 2022 – Present

ML/AI

Various

- 3rd place in NeurIPS2024's NMMO reinforcement learning track, showcasing advanced MARL and optimization skills.
- Won 3rd place in AMAZON's SUADD23 Mono Depth Estimation on drone images and presented my work at the German Conference on Pattern Recognition (GCPR23).
- 2nd place in NVIDIA's object tracking and detection competition, highlighting proficiency in tracking and real-time processing.
- 3rd place in Responsible AI's event for NN/ML model attacks and defence mechanisms.

Projects/Publications

Neural MMO - Multi-Agent Reinforcement Learning | *PPO, MARL, LLM, CTCE*

Jan 2024

- Secured 3rd place in NeurIPS24 MARL competition. Implemented advanced communication mechanisms and integrated Salesforce-Xgen-7B LLM, leading to improvement in agent collaboration and reduction in training convergence time with help of cross attention, LSTM, Centralized training.

Enhancing Local Business Intelligence | *RAG, Information Retrieval, Multi-Modal AI, Quantization*

June 2024

- Developed a location-based recommendation and Q&A system using LLaMA-3, multi-modal RAG, and CLIP embeddings. The system improved local business engagement and reduced response time by 40% using compression/pruning, demonstrating full-stack development skills and impact in real-world scenarios.

AI-Driven Generic Recruitment Platform | *RAG, NLP, LLMs, FastAPI, LLaMAIndex*

Current

- Developed AI services to streamline recruitment processes, including automated JD creation, job recommendation, and candidate ranking using RAG techniques, resulting in increase in recruitment efficiency and reduction in hiring time.