

SAIDINESH POLA

Ecity, Bangalore

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

Experience

Infosys

June 2021 – Current

Specialist Programmer L3 (AI/ML)

Bangalore, India

- Led the development of Gen-AI recruitment platform APIs, implementing NLP techniques including Named Entity Recognition using fine-tuned BERT models for resume parsing and candidate recommendation algorithms.
- Implemented advanced computer vision and NLP solutions, including multimodal image retrieval systems and chatbot functionality, showcasing expertise in intelligent automation technologies.
- Designed automated ML pipelines covering data preprocessing, synthetic data generation (RAFT), model training, and evaluation, enhancing efficiency in recruitment and scholarship platforms using closed/open LLMs.
- Developed frontend modules using TypeScript and Angular, integrating with backend systems for Scholarship Platform.
- Actively participated in top AI hackathons (NeurIPS, CVPR, KDD), securing wins in NeurIPS24, CVPR24, and GCPR23, demonstrating cutting-edge expertise in AI/ML research and applications.

AICrowd

April 2023 – July 2023

Research Fellowship

Remote

- Developed a CV model for WHO's schistosomiasis prevention, achieving 52.5 Macro F1 score, and built an accompanying web app, demonstrating ability to deliver end-to-end ML data collection pipeline with real-world impact.
- Contributed to a Cybersecurity Risk Assessment RAG application, showcasing experience with NLP, LLMs, and their practical applications in security contexts.

Education

Indian Institute of Technology, IIT Guwahati

June 2019 – May 2021

Communication Engineering, MTech

GPA 7.47

Paper: Implementation of a new order-preserving Lossless Data Compression scheme to reduce overhead in a Novel Interdependent Joint Source and Channel Coding in wireless sensor networks

National Institute of Technology, NIT Raipur

June 2014 – May 2018

Electronics and Communication Engineering, BTech

Technical Skills

Programming: Python, TypeScript, JavaScript, SQL, C++

ML/AI Frameworks: PyTorch, TensorFlow, Transformers, RAG, LlamaIndex, Rllib

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

Cloud & Tools: AWS, GCP, Git, Jupyter, WANDB, Power BI

Specializations: Machine Learning, NLP, Computer Vision, Reinforcement Learning, OCR

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

- 2nd runner-up in CVPR2024's Vision Language Modeling few-shot anomaly detection challenge, demonstrating expertise in computer vision and NLP integration.
- 3rd place in NeurIPS2024's NMMO reinforcement learning track, showcasing advanced AI and optimization skills.
- 2nd place in NVIDIA's object tracking and detection competition, highlighting proficiency in computer vision and real-time processing.

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, demonstrating expertise in complex AI systems and optimization.

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. Created workflows and user flows to optimize the platform’s efficiency.

Enhancing Local Business Intelligence | *RAG, Information Retrieval, Multi-Modal AI* June 2024

- Developed a location-based recommendation and Q&A system using LLaMA-3, multi-modal RAG, and CLIP embeddings. Implemented FastAPI backend, Streamlit frontend, demonstrating full-stack development skills.