Samir Sengupta

New York, NY, USA · +15513591228 · samir843301003@gmail.com

github.com/SamirSengupta · linkedin.com/in/samirsengupta · esp1420.github.io/ClientESP-0/samir

Education

M.S. in Data Science, Saint Peter's University, NJ, USA, GPA: 3.9/4.0
Sep 2024 – Dec 2025 Focus: LLMs and Multi-Modal AI
Relevant Coursework: Deep RL, Advanced Neural Networks, MLOps, Computer Vision, NLP

B.S. in Data Science, University of Mumbai, India, GPA: 3.8/4.0
Aug 2020 – Apr 2023 Capstone: Multi-Agent Reinforcement Learning
Relevant Coursework: Machine Learning, Deep Learning, Statistics, Data Structures, Optimization

Professional Experience

AI Researcher, Nov 2024 – Mar 2025
Saint Peter's University, NJ, USA

- Led LLM fine-tuning (LoRA, QLoRA) achieving 45% task-specific performance improvement
- Architected RAG pipelines with Pinecone vector DBs improving contextual accuracy by 60%
- Built distributed LLM training reducing inference latency by 35%
- Mentored 3 junior researchers: published findings at AI conferences

Software Developer, Aug 2023 – Jul 2024
Synradar, Mumbai, India

- Developed ML-based intrusion detection systems improving detection accuracy by 40% and reducing false positives by 25%
- Built automated threat-hunting pipelines processing 10M+ events/day
- Integrated LLaMA 3.1 with security knowledge base; evaluation speed up 60%
- Created real-time dashboards using React + D3 reducing incident response time 35%
- Led 5 engineers implementing full MLOps deployment pipelines

Data Scientist, Jan 2022 – Jul 2023
Neural Thread, Mumbai, India

- Designed ML pipelines improving forecast accuracy 25%, generating $10K revenue
- Developed generative AI chatbots increasing engagement 35%, reducing content costs 40%
- Optimized ML models via compression/quantization reducing costs 50% with 99.5% accuracy
- Implemented A/B testing improving user retention 20%

Key Projects

- **AI Finance Copilot**: Built Python/SQL system with anomaly detection, forecasting, and interactive dashboards for personal finance.
- **NotebookAI**: RAG-powered document assistant using LangChain, FAISS/Postgres, FastAPI, and Streamlit with Plotly visuals.
- **LLM Agent System**: Autonomous modular agent system using Gradio and Playwright; containerized with Docker and optimized inference.

Publications

Efficient Fine-tuning of LLMs for Domain Applications (Medium); Multi-Modal Fusion Networks for Medical Image Analysis (Saint Peter's AI Conf 2025); 3 conference presentations.

Certifications: TensorFlow Developer Professional, AWS Certified ML Specialty, Google Cloud ML Engineer, Azure AI Engineer, NVIDIA Deep Learning Institute

Leadership & Impact

Led 15+ cross-functional ML projects and mentored 8 junior engineers. Delivered over $50K in business value through AI automation, reduced operational costs by 40%, and improved customer satisfaction by 25%. Active open-source contributor and technical blogger (5K+ monthly readers). Regular speaker at AI/ML meetups.