

# Samir Sengupta

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

GitHub: [github.com/SamirSengupta](https://github.com/SamirSengupta) • LinkedIn: [linkedin.com/in/samirsengupta](https://linkedin.com/in/samirsengupta) •

Portfolio: [esp1420.github.io/ClientESP-0/samir](https://esp1420.github.io/ClientESP-0/samir)

## PROFESSIONAL SUMMARY

---

AI Engineer and Data Scientist with 4+ years of experience building enterprise scale Machine Learning solutions, Large Language Models, and production Generative AI systems. Specialized in computer vision, Natural Language Processing, GenAI, and Neural Networks with proven track record of delivering \$50K+ business impact through predictive analytics and automation. Currently pursuing MS in Data Science. Authorized to work in US.

## TECHNICAL SKILLS

---

- Programming: Python, SQL, R, Scala, Java, C++, JavaScript, Bash
- Machine Learning/Deep Learning: TensorFlow, PyTorch, Keras, XGBoost, LightGBM, scikit-learn, Hugging Face
- GenAI/LLMs: GPT, LLaMA, BERT, T5, Stable Diffusion, LoRA, QLoRA, RAG, LangChain, OpenAI API, N8N
- Computer Vision: OpenCV, YOLO, R-CNN, U-Net, Object Detection, Image Segmentation, OCR
- Data Engineering: Spark, Kafka, Airflow, Databricks, Snowflake, dbt, ETL/ELT
- Cloud/MLOps: AWS (SageMaker, EC2, S3), GCP (Vertex AI), Docker, Kubernetes, MLflow, CI/CD
- Databases: PostgreSQL, MySQL, MongoDB, Redis, Elasticsearch, Neo4j, ClickHouse
- Analytics/Visualization: Tableau, Power BI, Plotly, A/B Testing, Statistical Analysis, Time Series

## 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

### AI Solutions Architect — 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-Powered Personal Finance Copilot:** Built Python-based finance system integrating ML, SQL, and interactive dashboards. Implemented anomaly detection, predictive forecasting, and real-time visual insights.
- **NotebookAI: RAG-Powered Document Assistant:** Developed RAG pipeline for document upload, semantic search, and contextual QA. Used LangChain, FAISS/Postgres, FastAPI backend, Streamlit frontend with interactive Plotly visuals.
- **LLM-Powered Autonomous Agent System:** Designed modular agent architecture for autonomous research and automation using Gradio and Playwright. Deployed scalable containerized system via Docker with optimized inference pipelines.

## EDUCATION

---

**M.S. in Data Science | Saint Peter's University, NJ, USA | Sep 2024 – Dec 2025**

GPA: 3.9/4.0 | 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 | Aug 2020 – Apr 2023**

GPA: 3.8/4.0 | Capstone: Multi-Agent Reinforcement Learning

Relevant Coursework: Machine Learning, Deep Learning, Statistics, Data Structures, Optimization

## PUBLICATIONS CERTIFICATIONS

---

- Efficient Fine-tuning of LLMs for Domain Applications (Medium)
- Multi-Modal Fusion Networks for Medical Image Analysis — Saint Peter's AI Conference 2025
- 3 conference presentations
- Certifications: TensorFlow Professional, AWS ML Specialty, Google Cloud ML Engineer, Azure AI Engineer, NVIDIA DLI

## LEADERSHIP IMPACT

---

Led 15+ cross-functional ML projects and mentored 8 junior engineers. Delivered \$50K+ value via AI automation; reduced costs 40% and improved customer satisfaction 25%. Active open-source contributor; technical blog with 5K+ monthly readers; regular AI/ML speaker.