

Sriram Iyengar

Data Scientist | Machine Learning Engineer
GenAI Systems | Healthcare and Fintech

sriramkiyengar@gmail.com

linkedin.com/in/sriramiyengar2001/

SUMMARY

Data Scientist with 2+ years building predictive models, anomaly detection systems, and GenAI/LLM solutions across healthcare and fintech. Delivered \$20M+ in client impact through statistical modeling and deep learning on large-scale claims data. AWS Certified Cloud Practitioner; skilled in translating complex analysis into stakeholder-ready insights.

EDUCATION

The Pennsylvania State University

B.S. in Data Science (Computation and AI focus)

2 x Dean's List Recipient | Lecture Assistant, Department of Data Science

Pennsylvania, USA

Class of 2023

WORK EXPERIENCE

Independent AI Consultant

November 2025 - Present

Qapita | F&D Partners

Bangalore, India

- Engineered financial analytics automation pipeline for a deregulated energy advisory firm; combined agentic browsing for real-time data extraction with custom Python workflows, cutting supplier evaluation time from 3 hours to minutes.
- Built multi-jurisdiction document automation pipeline processing 10,000+ records across 15+ regulatory contexts; architected multi-API orchestration layer with structured output validation, reducing report generation from 5 hours to 5 minutes (98% reduction). Dockerized and deployed to production.
- Engineered LLM-based data migration pipeline with deterministic fallbacks for non-standard field parsing; maintained zero data loss across full dataset transfer, cutting new client onboarding from days to hours.
- Instrumented 200+ analytics events across 5 core product workflows using Mixpanel; defined engagement metrics, built usage dashboards, and surfaced insights that directly informed product roadmap decisions for senior leadership.

ML Engineer – Growth

July 2025 - October 2025

Mason

Pennsylvania, USA

- Built a Python ETL pipeline aggregating multi-source social media data for sales stakeholders; automated collection and normalization, improving lead-gen productivity by 50%.
- Developed an end-to-end content analysis and personalization pipeline using Google Gemini Vision Models, reducing internal manual workload by 75%.

Junior Data Scientist - Healthcare Analytics

November 2023 - March 2025

Certilytics Inc.

Pennsylvania, USA

- Delivered up to \$20M in contract savings for Fortune 500 health plans via deep learning anomaly detection on claims data at scale.
- Reduced model input dimensionality from 320 to 17 features (95% reduction) via unsupervised feature extraction on AHRQ Social Determinants of Health data; improved production accuracy and passed independent scientific review.
- Automated client deployment recalibration pipeline; reduced new client setup time by 71% through scripted retraining and validation workflows.
- Refactored Scala and SQL data processing pipeline, resolved systemic inefficiencies that reduced error rate by 30% and cut pipeline runtime by 40%.

KEY PROJECTS

Multi-Agent RAG Evaluator | Python, Google ADK, Vertex AI, Pinecone, Streamlit

- Architected a 4-agent RAG system on Google ADK and Vertex AI with Gemini-based reranking and full observability; validated output quality using a dual RAGAS + LLM-as-a-judge evaluation framework on a 20-question domain-specific test set via DeepEval

LiftVision: End-to-End CV Inference + Feature Engineering Pipeline

- Built a workout biomechanics analysis system using MediaPipe pose landmarking; extracted joint angles, bar velocity progression, and bar path features across 4 barbell movements; deployed as a versioned python service with automated technique-flagging and body proportion analysis.

TECHNICAL SKILLS

- Languages:** Python, SQL, Apache Spark/Scala, R.
- ML/AI:** PyTorch, scikit-learn, Deep Learning, Statistical Analysis, Predictive Modeling, Data Modeling, RAG, Anomaly Detection, Computer Vision (MediaPipe, CNNs), Model Calibration, MLflow, Google ADK, Ethical AI.
- GenAI & LLMs:** LangChain, LangGraph, OpenAI API, GCP (Vertex AI, Gemini API), RAG, Multi-Agent Systems, LLM Fine-tuning (LoRA/QLoRA), Prompt Engineering, Vector DBs (Pinecone, FAISS), Agentic AI, RAGAS, LangSmith, DeepEval.
- Data & Engineering:** Pandas, NumPy, FastAPI, ETL Pipelines, PyMC, MixPanel, Tableau, Postman, Alteryx, Github.
- MLOps & Infra:** Docker, AWS (Certified), Vertex AI, CI/CD for ML, Model Monitoring.