Laxmi Narasimha Hari Yelesetty

+91 99161 12881 | ylnharimailme@gmail.com | LinkedIn | GitHub | Portfolio

Lead AI/ML Engineer & Data Scientist | Production ML & GenAI | Distributed Systems | GCP · Azure · GPU Computing

Lead AI and ML engineer and data scientist with **10+ years** of experience, building production AI/ML/LLM systems. Leads end-to-end AI delivery, combining hands-on model development with technical leadership to turn business problems into scalable, high-throughput, low-latency production systems. Experience spans predictive modelling, forecasting, operations research, computer vision, and generative AI across retail, cybersecurity, manufacturing, insurance, and supply-chain optimization.

Skills

- **Agentic AI & LLMs:** Google ADK; Gemini / Vertex AI; Tool Calling; Agent Evaluation; LoRA / QLoRA; OpenAI-Compatible Serving
- **Modelling & Analytical Methods:** Classical ML; Deep Learning; Computer Vision; NLP; Forecasting; Feature Engineering; Operations Research; Autoencoders
- **Frameworks & APIs:** scikit-learn; XGBoost; PyTorch; TensorFlow; Hugging Face / PEFT; Unsloth; vLLM; FastAPI; ONNX Runtime; OR-Tools
- **GPU & Distributed Computing:** Multi-GPU Training; RAPIDS cuDF / CuPy
- **Model Lifecycle:** MLflow; Evaluation / Registry; Data / Model Quality Gates; Reproducible Training; Vertex AI / Azure ML Pipelines; Feature Store; Semantic Retrieval
- **Cloud & Deployment:** GCP / Azure; GKE / Kubernetes; Docker; Cloud Run / Functions; Azure Functions / Logic Apps; Helm / Istio
- **Reliability & Reporting:** CI/CD; Cloud Build; Argo CD / GitOps; Prometheus / Grafana; OpenTelemetry / Cloud Trace; GitHub Actions / pytest; Load Testing; Streamlit / Power BI
- **Languages, Data & Storage:** Python; SQL; Beam / Dataflow; Pub/Sub; BigQuery; Redis; GCS / Azure Blob; Pandas / NumPy; Dask; Git / Linux

Experience

Lead Artificial Intelligence Engineer — **Best Buy** | Jul 2025 - Present

Lead AI/ML architecture and end-to-end delivery across LLMs, search, recommendations, fraud and cybersecurity.

- Designed and built an **autonomous root-cause analysis agent** as sole developer with **Google ADK, Gemini on Vertex AI and Python**. Investigates production incidents using metrics, logs, traces, Kubernetes state and source code, with scoped tool access, read-only guardrails and evidence-backed reports; reduces investigation time from **hours to minutes**.
- Productionized a reusable **LLM fine-tuning framework** with Unsloth, PEFT, Hugging Face Transformers and Accelerate, supporting LoRA/QLoRA, distributed training, task-specific adapters and automated evaluation.
- Deployed **vLLM inference on GKE** with OpenAI-compatible endpoints, HPA autoscaling, Istio routing and Prometheus/Grafana monitoring, with **estimated annual API-cost savings of ~$300K**.
- Lead ML engineering for **sponsored-product recommendations handling nearly 3 million requests/day**: automated feature/training pipelines, model-quality gates, semantic retrieval and serving on GKE across two regions. Set architecture and release standards with data science, application, cloud and reliability teams; introduced reusable AI coding-agent skills and automated checks.
- Built retail fraud feature pipelines at **40,000 events/sec at peak** using Beam/Dataflow and Vertex AI Feature Store. Improved holiday score coverage by **~20%** and deployed XGBoost through FastAPI on Kubernetes; the model's contribution saves **US$1.5M/year**.

Lead ML Engineer - ML Engineering & Operations — **Wipro** | Aug 2023 - Jul 2025

- Engineered **AWS account digital fingerprinting** on NVIDIA DGX (**8 x A100**) using PyTorch autoencoders and RAPIDS cuDF/CuPy on **billion-row datasets**, with continuous retraining, MLflow and Grafana alerts.
- Scaled GPU-accelerated **phishing detection to 2,400 emails/sec (2x throughput)**; integrated analyst feedback and Microsoft Sentinel for detection and quarantine.
- Automated ML/OCR-based **abuse-inbox triage, doubling case volume**. Mentored engineers, coordinated client/NVIDIA delivery, screened **300+ resumes** and conducted **50+ technical interviews**.

Senior Data Scientist - Data Science / Azure ML **Accenture** | May 2021 - Sep 2023

- Developed **product-quality prediction for BamaGruppen** with fully automated training and serving using Logic Apps and Azure ML platform services, including Azure ML pipelines, model registry and container deployments.
- Increased truck inspections from **130 to 160/day**, defect detection by **17%** and customer approval by **20%**, delivering **$2.5M annual savings**.
- Replaced manual Excel planning with automated last-mile routing and transport optimization using OR-Tools and mixed-integer models, integrating Azure Functions, Logic Apps and RPA to enable rapid replanning and improve delivery efficiency.
- Modelled **supply-chain and distribution-hub scenarios** across cost, freshness, carbon and warehouse volumes using OR-Tools on Azure ML with Power BI reporting.
- Led cross-functional delivery, mentored data scientists and presented network-design recommendations to Norwegian stakeholders.

Data Scientist **GD Research Centre Pvt Ltd** | May 2020 - May 2021

- Developed **forecasts and scenario analysis for GlobalData** across **15+ industry verticals** with forecasting and feature-engineering pipelines, informing market intelligence and resource allocation.

Deep Learning Engineer **Tata Consultancy Services** | Jul 2019 - Mar 2020

- Developed **semiconductor defect detection** with TensorFlow / Keras DenseNet on a GCP multi-GPU cluster. Distributed training on **~1 million images in ~3 hours** achieved **90% classification accuracy**.
- Ingested SQL Server data and served inference with Docker / TensorFlow Serving. Deployed the end-to-end system at Amkor, South Korea, and trained the client team.

Senior Software Engineer - ML **Infosys** | May 2016 - Jul 2019

- Implemented **opportunity win/lose prediction** with logistic-regression scoring and data pipelines for sales prioritization; integrated SQL Server Integration Services, SSAS and Power BI dashboards.
- Delivered tested **insurance ID-card workflows** using COBOL, JCL, DB2, VSAM and CICS; handled impact analysis, design and support, and shared knowledge with the team.

Selected Recognition

Best Buy Talent Quotient (2026) - cybersecurity and fraud delivery.

Best Buy client AML Hall of Fame (2024), during the Wipro engagement - DGX framework delivery and offshore mentoring.

Accenture Delivery Excellence; Kudos x2.

Certifications

- Google Cloud Certified: Generative AI Leader | Microsoft Certified: Azure AI Engineer Associate
- Microsoft Certified: Azure Data Scientist Associate | Microsoft Certified: Power BI Data Analyst Associate
- Oracle Cloud Infrastructure 2024 Generative AI Certified Professional

Education

B.E. Electrical & Electronics **Andhra University, Visakhapatnam** | 2012 - 2016 · 80%