Theyaneshwaran Jayaprakash

+1 (812) 822-7804 | theyaneshj17@gmail.com | linkedin.com/in/theyaneshj17 | github.com/theyaneshj17/

Education

Indiana University Bloomington August 2024 - May 2026
Master of Science, Data Science; GPA: 4.00 *Bloomington, United States*

College of Engineering Guindy, Anna University August 2016 - April 2020
Bachelor of Engineering, Electronics and Communication Engineering; GPA 3.70 *Chennai, India*

Research Experience

ML Research Assistant September 2024 - Present
IU School of Medicine *Indiana, USA*

- Developed a self-supervised autoencoder in PyTorch, guided by Gaussian Mixture Model clustering for classification tasks, achieving **92% accuracy in Alzheimer's subtype discovery** (First Author, AAIC 2025; Fellowship Recipient).
- Validated identified subtypes via multidimensional analysis across neuroimaging and behavioral data domains, employing ANOVA, hypothesis testing, and **Bayesian MCMC inference** to quantify statistical significance.

Professional Experience

AI Engineer Intern June 2025 - Present
FocusKPI, Inc *California, USA*

- Architected and deployed a scalable **Model Context Protocol (MCP) client-server system** on AWS ECS, processing 75,000+ FDA medical device records with RAG and BGE embeddings for semantic search across 200 API fields.
- Engineered a real-time **multi-modal AI agent** with desktop overlay integration, achieving sub-5 second latency for speech transcription and visual analysis using Whisper and GPT-4V, to generate automated summaries and recommendations.
- Orchestrated **multi-agent workflows** using LangChain and LangGraph for context management across audio, visual, and text data streams.

Senior Data Engineer March 2023 - August 2024
UnitedHealth Group *Chennai, India*

- Optimized ELT efficiency by parallelizing workflows and **tuning SQL performance**, reducing monthly processing time from **5 to 2 days** and cutting compute costs by 50% (6K to 3K Data Warehouse Units).
- Migrated **50TB+ data** from Synapse to Snowflake using PySpark and SQL with ADF for phased migration with continuous sync, implementing **dbt transformation** layer for analytics.
- Developed **Power BI semantic models** in Microsoft Fabric, optimizing Direct Lake/Query modes with row-level security. Built 10+ interactive dashboards, **enhancing performance by 67%** via fine-tuned DAX queries.

Data Engineer October 2020 - February 2023
UnitedHealth Group *Hyderabad, India*

- Integrated **25+ data sources** (Parquet, JSON, CSV, Delta) from Oracle, Teradata, and Databricks into Azure Synapse using a **Medallion architecture**, improving ELT scalability.
- Built a FHIR streaming pipeline on Databricks with Delta Lake and PySpark, processing **5,000+ records/min** from Azure Event Hubs using optimized micro-batch triggers and partitioning.
- Migrated APIs from Azure VMs to Container Instances, enabling cross-cloud Snowflake queries, **reducing costs by 60% ($30K annually)** and improving execution time from 60s to 15s using C# and SQL optimizations.
- Automated CI/CD for 7 repositories with GitHub Actions, **halving release cycles** and improving deployment reliability.

Skills

Programming & Technologies: Python, R, C++, SQL, Git, Docker, Kubernetes, Hadoop, PySpark, A/B Testing
Data Engineering & Cloud: Azure (Synapse, ADF), AWS (S3, Lambda, Redshift), Databricks, Snowflake, dbt, Airflow
Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Tableau, Power BI, ggplot2, dplyr
Machine Learning & AI: PyTorch, TensorFlow, Scikit-learn, MLflow, RAG, Transformers, Causal Inference

Projects

IntelliFlow Enterprise AI Orchestration System (GitHub) | *LLM, Neo4j, FastAPI, MCP* August 2025

- Architected LLM-powered workflow orchestration with intent classification across 4 MCP servers, constructing Neo4j knowledge graph with **50K+ entities** enabling **multi-hop reasoning** for IT, data, and research automation.
- Developed secure **text-to-SQL engine** with parameterized query validation, real-time ServiceNow integration, and **GraphRAG**-based research capabilities for enterprise query processing.