Aneesh Jayan Prabhu

(602) 768-6622 | aneeshjayan11@gmail.com | linkedin.com/in/aneeshjayan | github.com/aneeshjayan

SKILLS

Languages: Python, SQL, R, C, C++, MATLAB

Frameworks & Tools: FastAPI, Matplotlib, Seaborn, TensorFlow, PyTorch, OpenCV, Scikit-Learn, Pandas, NumPy, MLflow, PowerShell, Microsoft Graph API, Exchange, Outlook distribution lists, Remedy, Power Automate

Databases: MySQL, PostgreSQL, Google Cloud SQL, Neo4j, MongoDB

Cloud & DevOps: AWS (Lambda, S3, SageMaker),Databricks (Jobs/Workflows, Feature Store), Azure DevOps, Docker, Git, CI/CD pipelines, Github Actions, MLOps, Dataiku

Data Engineering: ETL, Data Pipelines, Data Ingestion, Data Cleaning & Validation, Workflow Orchestration (Airflow), Batch/Streaming (Kafka), Distributed Processing (PySpark, Hadoop), Grafana

Data Science & ML: Regression, Classification, ime-Series Forecasting, Clustering, Neural Networks, Feature Engineering, EDA, Hypothesis Testing, Power BI

LLM & GenAI: LangChain, LangGraph, Model Context Protocol, Vector embeddings, Semantic search, Prompt Engineering, LoRA/QLoRA

EXPERIENCE

Data Science Intern   May 2025 - December 2025

Wolters Kluwer - Legal & Regulatory Division   *New York, USA*

- Engineered an AI-driven reporting agent using **LangGraph** pipelines for generating semantic dashboard insights via conversational UI, boosting reporting efficiency by **22%**.
- Reduced manual escalations by **41%** by deploying **FastAPI** microservices integrated with **PySpark + MongoDB**-backed workflow data and **CI/CD** via **Azure DevOps**.
- Crafted a **RAG preprocessing pipeline** for unstructured legal documents, applying flowchart and table transcription strategies to enhance retrieval quality.
- Benchmarked **transformer models (BERT, RoBERTa, T5)** within **RAG pipelines**, improving factual accuracy by **85%**.
- Presented **KPI dashboards** to legal/business stakeholders, translating model tradeoffs into actions.
- Increased ingestion reliability by **95%** and reduced latency by **42%** by automating SMTP to OneDrive ingestion and email-triggered workflows using Microsoft Graph API, with production monitoring/alerts.
- Participated in design reviews, sprint planning, and production monitoring, aligning system performance with product goals.

Research Intern - Biomedical and Neuroinformatics   May 2023 - May 2024

Centre for Cyber-Physical Systems, VIT   *Chennai, India*

- Engineered **EEG and fMRI** preprocessing pipelines using **Spatio-Spectral Decomposition (SSD)** in MATLAB and custom **deep learning architectures** for neuro-signal analysis.
- Built a hybrid **deep learning-quantum framework** for Autism detection from fMRI, integrating **Swin Transformers, CNNs**, and **Quantum SVM/QNN** models.
- Achieved **98.17% (ABIDE I)** and **96.2% (ABIDE II)** accuracy with **25% reduction** in computation time, demonstrating scalable AI-driven diagnostics.

ACADEMIC PROJECTS

Optimal-SLM: Reasoning and Prompt Optimization Framework   August 2025

- Built dual-agent system using **Qwen** and **Phi**, **fine-tuned** with **PEFT** on Alpaca and OpenOrca datasets, achieving **76% token reduction** while preserving meaning.
- Deployed via **Docker and CI/CD** on **AWS SageMaker**, ensuring **scalability, security, and compliance**.
- Designed **Streamlit dashboard** displaying intent analysis, optimized prompts, and metrics (quality score greater than **0.8**).
- Applied **4-bit quantization and LoRA** for efficient fine-tuning in low-resource (**6GB VRAM**) environments.
- Extended to chat + voice interface for real-time optimization of voice-to-text prompts.

FocusMate − AI Co-Pilot for Executive Function (spark Challenge Hackathon)   November 2025

- Built an ADHD-focused co-pilot that converts **Gmail and Google Calendar** + voice notes into prioritized tasks, email summaries, and **time-blocked daily plans** with **90-minute break scheduling** to reduce overwhelm and time-blindness.
- Engineered **2 FastAPI microservices** (ports **8000/8001**) with Google **OAuth2** and **Gmail/Calendar APIs** to ingest/cache emails, classify **action vs. informational**, extract due-dated next steps, and enable natural-language inbox search.
- Delivered a full-stack MVP with a **React (Vite)** planning dashboard, optional **Expo** mobile client, and voice-to-task pipeline; exposed **6+ REST endpoints** (refresh, search, summaries, STT) for end-to-end triage and planning.

TrustMedAI: Medical Conversational Agent for Type-2 Diabetes   December 2025

- Built semantic retrieval system processing **500+ forum threads** and **16,000 lines** of clinical guidelines from **ADA, Mayo Clinic**, and **NIH** using regex normalization, deduplication, and **Key Recurring Themes (KRT)** identification via **MiniLM embeddings**.
- Developed production **RAG pipeline** with **FAISS vector database** achieving Content Precision **0.950**, Recall **0.920**, Faithfulness **0.970**, KRT-F1 **0.850**, ROUGE-L **0.850**, METEOR **0.780**, BERTScore **0.930**, and Sentence Similarity **0.888**.
- Designed **React-based** multimodal interface with **speech-to-text** and **text-to-speech** for accessible patient education featuring citation-backed responses.

EDUCATION

Arizona State University   Tempe, Arizona, USA

MS in Data Science, Analytics and Engineering   *May 2026*

- Coursework: Statistics for Data Analysts, Data Processing At Scale, Statistical Machine Learning, Data Mining, Knowledge Representation and Reasoning, Computing for Data-Driven Optimization.

Vellore Institute of Technology (VIT)   Chennai, India

B.Tech in Electronics and Communication Engineering   *May 2024*