Aparna Jayakumar Nair

DATA SCIENTIST, DATA ANALYST

Las Cruces, NM — 575-650-7378
aparnajj@nmsu.edu / aparnajjn@gmail.com
LinkedIn: https://www.linkedin.com/in/aparna-jayakumar-nair/
GitHub: https://www.github.com/Aparna-J-Nair

Summary

- Experienced Machine Learning and NLP professional with 4+ years of expertise in designing and deploying AI-driven solutions using TensorFlow, Flask, and LLM models.
- Proficient in Python, SQL, R, and machine learning libraries for natural language processing, deep learning, LLMs, and data analytics.
- Strong research and industry background with impactful AI solutions across diverse domains.

Education

MS in Data Analytics,
New Mexico State University, USA — **Jan 2024 - Present**
GPA: 3.85/4.00

MTech in Data Science,
Amrita Vishwa Vidyapeetham, India — **Aug 2020 – Jul 2022**
GPA: 8.64/10.00

BTech in Electrical and Electronics Engineering,
NSS College of Engineering, India — **Aug 2015 – Aug 2019**
GPA: 7.27/10.00

Skills

Programming: Python, R, C, SQL, Julia

AI-ML Technologies/Frameworks: MATLAB, TensorFlow, Keras, Scikit-learn, NLTK, spaCy, Flask API, OCR, OpenCV, Generative AI(Gen AI), Databricks, Palantir, Pytorch, Langchain, Open AI api, ElasticIndex, ONNX, RAG, LlamaIndex.

Cloud platforms: Azure OpenAI Service, Google Vertex AI

Experience

Cummins Inc. - AI Intern — **May 2025 - Aug 2025**

- Developed a framework for clustering Voice of Customers (VOC) data to extract actionable insights.
- Built a Retrieval-Augmented Generation (RAG) application for a safety-focused chatbot using large language models (LLMs), earning 3rd place in an internal competition.
- Applied semantic mapping techniques to enhance data interpretation and model performance.

New Mexico State University - Research Assistant — **Jan 2024 - Present**

- Designed and implemented a Modbus TCP data pipeline that collects, logs, and decodes solar power plant data with real-time processing and configurable endianness for analytics.

- Collected and processed solar power system data via Modbus TCP, applying decoding, statistical analysis, and machine learning techniques for visualization and insights.
- Developed an algorithm to analyze shading effects in single-axis solar panels.

New Mexico State University - Teaching Assistant **Jan 2025 - Present**

- Conducting classes on Python programming and AI applications.

LTI Mindtree - Software Engineer (Data Science) **Aug 2022 – Dec 2023**

- Built predictive APIs for real-time text completion and data correlation using Generative AI models, reducing manual efforts by 40%.
- Developed an OCR-based system for extracting and structuring image content with 95% accuracy.
- Led a topic modeling project, and developed API for extracting key topics from unstructured data and fine-tuning an LLM to generate precise topic names for each cluster, improving interpretability.
- Led projects on image/audio-to-text conversion and correlated extracted data with SDLC.
- Designed a text completion API leveraging LLMs similar to Google's Smart Compose.
- Developed an API for Natural Language Generation (NLG) to automate text creation from structured data, ensuring coherence and contextual relevance.

L&T Infotech – Data Science Intern **Jan 2022 – Jul 2022**

- Worked on Knowledge Graphs and Entity Recognition in NLP.

University of Wyoming - Research Assistant **May 2021 – Jan 2022**

- Conducted gait analysis for cybersecurity using Machine Learning, leading to a Springer publication.

Publication

"Continuous Authentication Using Gait Patterns", Proceedings of the second International Conference on Signal and Data Processing,2023, Springer Nature Singapore, Page:447–459, ISBN: 978-981-99-1410-4.", Springer, 2023.

Professional Projects

Data-Based Analysis of Solar Plant - NMSU **Jan 2024**

- Designed and implemented a Modbus TCP data acquisition pipeline to collect real time solar power plant data for organized long-term monitoring.
- Developed a real-time data processing system that decodes raw Modbus register values (float32/float64, configurable endianness) into engineering units ensuring robustness with parallel processing and fault tolerance.
- Conducted data analysis for anomaly detection in solar plant operations using statistical and machine learning techniques.
- Developing real-time interactive dashboards for data visualization to enhance monitoring and decision-making processes.

Designing and Shading Analysis for AgriVoltaics- Research Assistant - NMSU **Jan 2024**

- Designed solar panels for 3 different locations, shading analysis of different types of solar trackers is conducted.
- Proposed algorithm and mathematical formula for the shading analysis of single-axis solar panels.
- It is specifically analyzed in collaboration with the agriculture department funded by DOE.

Smart Compose for Text Completion - LTI Mindtree **Aug 2023**

- Developed an API was built for live streaming of text from a Generative AI model like the Gmail Smart Compose feature.
- Improved text generation accuracy, leveraging fine-tuned AI models for high-quality predictions.
- Designed and implemented the system using RESTful API and HTML for seamless integration with web-based applications.
- Optimized response latency, ensuring faster text predictions and enhanced user experience.

Multimodal Data Extraction & SDLC Correlation - LTI Mindtree **May - Jul 2023**

- Developed APIs to extract and process image, audio, and video content using OCR, LLMs, and AI models.
- Automated the correlation of extracted insights with SDLC data, enhancing decision-making and documentation.

Rule-Based Natural Language Generation - LTI Mindtree **Mar 2023**

- Developed a rule-based system to generate dynamic visualizations and GPT-driven text insights based on user queries.
- Integrated pre-trained GPT models to enhance interpretability by providing contextual text explanations alongside output plots.
- Optimized query processing and automated data-to-text conversion, enabling seamless integration of visual analytics and text summaries.

Topic Modeling Using GPT - LTI Mindtree **Dec 2022**

- Developed an API to extract key topics from unstructured data using GPT, enabling automated topic identification.
- Fine-tuned an LLM model to generate precise and meaningful topic names for each cluster, improving interpretability.
- Optimized the topic extraction pipeline with advanced prompt engineering and clustering techniques for better accuracy.

Gait Analysis - University of Wyoming **Jun 2022**

- Developed a novel attack system on motion sensor data by leveraging correlations between motion sensor readings and gait video analysis.
- Explored security vulnerabilities in gait authentication systems, demonstrating potential risks in motion sensor-based authentication.
- Published findings in a research paper, presenting insights into adversarial attacks on gait-based authentication at an international conference.

Academic Projects - **Amrita Vishwa Vidyapeetham**

- **Question Answering Using BERT** **Dec 2022**
- **Automatic Speech Recognition Using HuBERT** **Dec 2022**
- **Underwater Acoustic Signal Extraction and Analysis** **Sept 2022**
- **Detecting depression in Tweets using Bayes Theorem** **May 2022**
- **Classifying movies using a deep neural network** **May 2021**
- **CNN based Speech emotion recognition** **Dec 2020**

Achievements

- Google Cloud Certified Generative AI Leader [Certificate]
- Won third place in IT Intern Challenge competition at Cummins.
- First prize in the WERC Environmental Design Contest 2024 for load optimization.
- Best project of the Electrical Engineering department for undergraduate project.

Certifications

- Generative AI Explorer - Vertex AI
- Gen AI Agents: Transform your organization (Google)
- Generative AI Leader Certification (Google Cloud)
- Data Analysis with Python (IBM, Coursera)
- SQL for Data Science (IBM, Coursera)
- Neural Networks & Deep Learning (Coursera)