Aparna Jayakumar Nair

aparnajj@nmsu.edu aparnajjn@gmail.com +1 (575) 650-7378 linkedin.com/in/aparna-jayakumar-nair
github.com/Aparna-J-Nair https://aparna-jayakumar-nair.vercel.app/

Education

New Mexico State University *Las Cruces, USA*
MS in Data Analytics, GPA: 3.85/4.0 *Jan 2024 - May 2026*
Amrita Vishwa Vidyapeetham *India*
MTech in Data Science, GPA: 8.64/10.0 *2020 – 2022*
NSS College of Engineering *India*
BTech in Electrical & Electronics Engineering, GPA: 7.27/10.0 *2015 – 2019*

Experience

Cummins Inc. *USA*
AI Intern *May 2025 – Aug 2025*

- Improved accessibility of safety procedures by 95% by developing a **multimodal data**(voice and text) **Retrieval Augmented Generation (RAG)** chatbot that provided sub-second, context-aware responses using **Azure OpenAI Service**, **Langchain**, **GPT-4o**, **Whisper**, and a **FAISS**-indexed knowledge base.
- Influenced data-driven product decisions, resulting in a 15% increase in customer satisfaction, by developing a **Clustering** framework on the **Databricks** platform to extract actionable insights from Voice of Customer (VOC) data.
- Improved data consistency by 98% and reduced manual data mapping efforts by 40% by developing a semantic matching pipeline on **Palantir** and **Microsoft Azure Open AI** to map NIST data to descriptions and assign compliance rules.

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

- Led a team of 4 data scientists across 3 concurrent projects, delivering a suite of **Generative AI** APIs using **Flask Python web framework** that reduced manual data processing efforts by 40%.
- Enhanced project decision-making by 50% by architecting a multimodal data pipeline with 95% accuracy, using, **OpenCV**, **Tesseract OCR**, **PyPDF2**, and **OpenAI Whisper** for data extraction, and leveraging **FAISS**, **Elasticsearch**, and **Sentence Transformers** for vector search to perform data correlation with SDLC data.
- Increased user productivity by 25% by designing and deploying a text completion API, similar to **Google's Smart Compose**, using **fine-tuned LLMs**.
- Enhanced data interpretability by 30% and reduced manual curation effort by 97% by architecting a **Python REST API** using **Flask** for end-to-end **Topic Modeling**; the system integrated **BERTopic** (**Sentence-BERT**, **UMAP**, **HDBSCAN**) for clustering with **Hugging Face Transformers** (**GPT-Neo**, **Flan-T5**, **BART**) to generate precise topic labels via **Prompt Engineering**.
- Automated report generation, saving 10 hours per week, by creating an API for **Natural Language Generation** (NLG) to produce coherent text summaries from structured data.

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

- Improved information retrieval accuracy by 25% by designing and implementing a **Knowledge Graph** with an LLM-based **NER pipeline**, using **PostgreSQL** and **Centrality Theorems** to model and rank **entity relationships**.

Skills

Programming: Python, R, SQL, C, Julia
AI/ML Technologies: TensorFlow, PyTorch, Keras, Scikit-learn, NLTK, spaCy, OpenCV,LlamaIndex, ONNX
Cloud & Data Platforms: Google Vertex AI,

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.

Projects

Protego Safety Assistant Chatbot *Cummins Inc.*
Jul 2025

- Achieved 99% precision in safety information retrieval by designing a multimodal data (voice and text) **RAG** chatbot using **Azure OpenAI Service**,**OpenAI GPT-4o**, **Whisper**, and **LangChain**.
- Engineered a data ingestion and retrieval pipeline using **LangChain** to create a semantic knowledge base with **OpenAI Embeddings** and **FAISS** for efficient vector search.

– Enhanced user accessibility by developing a dual-mode interface processing both text and voice queries, utilizing **OpenAI Whisper** for real-time speech-to-text.

Gait Analysis for Cybersecurity *University of Wyoming*
2022

– Exposed a critical cybersecurity vulnerability, leading to a Springer publication, by developing a novel gait spoofing attack in **Python** that correlated video data from **OpenPose** (**Computer Vision**) with motion sensor features from **tsfresh**.
– Achieved >90% authentication accuracy using **Machine Learning** models (**SVM**, **Random Forest**, **Neural Networks**) with data balanced via **SMOTE** and features selected in **Weka**, before proving the attack's success by increasing the system's Equal Error Rate (EER).

Leadership & Achievements

– Led a team of 4 members and secured third place out of 20 teams in the IT Intern Challenge at Cummins by developing a high-impact AI solution.
– Achieved first prize in the WERC Environmental Design Contest 2024 out of 15 teams by designing a load optimization solution that demonstrated a 25% increase in efficiency.
– Led a team of 4 members and awarded Best Project in the Electrical Engineering department for an undergraduate project out of 37 teams that outperformed previous benchmarks by 80%.
– Led weekly **Python** & **AI** labs for 50+ students, resulting in a 20% improvement in average assignment scores.

Certifications

– Google Cloud Certified Generative AI Leader [Certificate]
– Generative AI Explorer - Vertex AI