Jishnu S Nair

Hyderabad, India | +91 8347508479

jishnunair.com | jishnunair.mec@gmail.com | linkedin.com/in/thejishnunair | github.com/thejishnunair

Summary

Skilled machine learning engineer with **6+** years of experience in solving real-world problems with ML and NLP. Proven track record of designing, developing, and deploying machine learning models that have a positive impact on businesses.

Publications

Apriel-Nemotron-15B-Thinker
arXiv:2508.10948 *August 2025*

DNR Bench: Benchmarking Over-Reasoning in Reasoning LLMs - AAAI 2026
arXiv:2503.15793 *March 2025*

Experience

ServiceNow **March 2024 – Present**
Senior Machine Learning Engineer *Hyderabad, India*

- **Reinforcement Learning for Agents**: Optimized large language models for agentic applications using advanced RL algorithms (GRPO, DAPO, PPO), achieving improved decision-making and task efficiency.
- **Agentic Orchestration**: Designed and implemented a multi-agent orchestration framework with **LangGraph** to coordinate complex tasks and automate enterprise-scale workflows.
- **Synthetic Data Generation Framework**: Built a graph-based framework leveraging **LangGraph** to generate high-quality synthetic datasets for Supervised Fine-Tuning (SFT) and Reinforcement Learning pipelines.
- **Reasoning Benchmark**: Developed a standardised benchmark to evaluate reasoning capabilities of LLMs across diverse tasks and scenarios.
- **Evaluation Framework**: Designed and developed a holistic evaluation framework using LangChain to analyse various capabilities and traits of LLMs by incorporating industry-standard datasets and metrics.

ServiceNow **September 2022 – February 2024**
Machine Learning Engineer *Hyderabad, India*

- **LLM Fine-tuning**: Fine-tuned and evaluated large language models like **StarCoder, StarCoderPlus, Llama 2** to perform tasks like text summarization, multi-turn Q&A and code generation specific to ServiceNow.
- **Inference Pipeline**: Developed an inference pipeline for large language models (LLMs) using **vLLM, TGI, and TensorRT**, incorporating model quantization and continuous batching
- **Dataset Framework**: Designed and developed a **LangChain**-based framework that uses open-source large language models and advanced **prompting** to generate synthetic data and perform various transformation.
- **Similarity Service**: Developed and architected an enterprise-level **Similarity Search** service using the state-of-the-art Transformer-based models, ONNX, **HNSW** and integrated NLP algorithms for accurate similarity analysis.
- **Clustering Service**: Built an efficient **KMeans Clustering** Service with prefiltering capabilities using **FAISS**, enabling hassle-free clustering of large datasets at an enterprise level.
- **Mentorship**: Provided mentorship and guidance to **two interns**, resulting in the successful completion of their internship projects.

ServiceNow **March 2022 – August 2022**
Software Engineer - AI/ML *Hyderabad, India*

- **ML Automation**: Developed and deployed ML-based automation to identify customer issues that could be resolved by ChatBot, deflecting **1.75M USD** worth of tickets from human resolution every quarter.

ServiceNow **July 2020 – February 2022**
Associate Software Engineer - AI/ML *Hyderabad, India*

- **Semantic Search**: Built and deployed semantic search service using the RoBERTa algorithm to retrieve relevant articles for customer incident tickets, resolving **1.35M USD** worth of tickets per year without human intervention.
- **Virtual Agent**: Created and maintained virtual agent topics using NLU (Natural Language Understanding), thereby improving customer experience and reducing costs.

DBS Partners **March 2019 – October 2019**
Machine Learning Intern *Remote, India*

- **Contract Analytics**: Developed and deployed NLP models to extract relevant features from employment contracts using custom datasets and advanced NLP algorithms, such as **RNN, LSTM**, and **RASA NLU.**

Technical Skills

Programming Languages: Python, Java, C, HTML/CSS, JavaScript, SQL.
Technologies/Frameworks: TensorFlow, Pytorch, Django, Flask, Git, Scikit-learn, Azure, Docker, ONNX, FAISS, EJML, LangChain, DeepSpeed, RASA NLU, Spacy, vLLM, TensorRT.
ML & NLP: Clustering, Classification, RNN, LSTM, BERT, RoBERTa, Optimization, LLMs, Prompting, Generative AI.
Team & Agile Collaboration: TDD, Agile development practices, Planning & estimation.

Education

Government Model Engineering College, Kochi **2016 – 2020**
Bachelor of Technology in Computer Science and Engineering *Kochi, India*
CGPA : 8.8/10

Patents

Platform Agnostic Scalable And High-performance Semantic Search Framework **April 2024**
US Patent Application No. 18/623,844 *Patent Filed*

Smart Persistence Of Model For Effective Predictions And Updates **April 2024**
US Patent Application No. 18/626,139 *Patent Filed*

Framework Neutral And Updatable Clustering Model **April 2024**
US Patent Application No. 18/426,810 *Patent Filed*

Projects

Web Browsing and Scraping using Natural Language Commands

- An end-to-end solution for voice-based web browsing and scraping developed using RASA Natural Language Understanding, Selenium and Chrome Extensions.

Growth Tracker

- An analytical dashboard that visualizes growth signals to help students track their progress and make informed decisions about their educational journey.

Awards and Recognition

ATG Llama Award: Recognized as a top contributor to the field of machine learning by ServiceNow, receiving the prestigious Llama Award **twice (2022 and 2023)** with a recognition ratio of just **14%**.

Positions of Responsibility

Secretary-General: Served as Secretary-General of MECMUN 2018, a techno-managerial fest organized by Government Model Engineering College, Kochi. Led the organization of the event and ensured its smooth execution.

Deputy Secretary-General: Served as Deputy Secretary-General of MECMUN 2017, a techno-managerial fest organized by Government Model Engineering College, Kochi. Assisted the Secretary-General in the organization of the event and played a key role in its success.