**Aneesh Mukkamala**

aneeshmukkamala@gmail.com | +91-7416829748 |[github.com](https://github.com/aneeshm11) |[linkedin.com](https://www.linkedin.com/in/aneesh-mukkamala-711203339/)

**EDUCATION**

**National Institute of Technology,** Andhra Pradesh  ***Nov 2022 – April 2026***

* Bachelor of Technology | Metallurgical and Materials Engineering ***CGPA****: 8.52 /10*

**TECHNICAL SKILLS**

**Languages:** Python, Java, CUDA, JavaScript, TypeScript, HTML, CSS

**Libraries/Frameworks:** PyTorch, TensorFlow, Huggingface, Transformers, LangChain, NumPy, SciPy, OpenCV, RoboFlow XLA, SimpleITK, Bootstrap, ReactJS, NodeJS, ExpressJS

**Cloud/Microservices:** AWS (EC2, SageMaker, DynamoDB, Lambda, Kinesis, Route53, CloudWatch, CloudTrail, RDS)

Boto3, Docker, MongoDB, Firebase, Git

**PROFESSIONAL EXPERIENCE**

**Resolute AI Software Limited | AI Engineer Intern *Nov 2024 – Feb 2025***

* Developed multi agent LLM based RAG system powered by Llama, Gemini, BERT variants and LangChain with persistent datastores, session states and chat history improving boot time and user interaction.
* Adopted robust pipelines using RAPTOR, RAG-Fusion, C-RAG for efficient indexing and data retrieval across 200 + documents with high accuracy and low processing and data retrieval time.
* Designed a data ingestion pipeline automating custom dataset creation followed by no-code LLM fine tuning interface using AWS SageMaker with end-to-end model deployment to multiple data hubs.
* Developed scalable knowledge-based RAG system using Cosmos DB for vector search and no-SQL integrations. Leveraged quantized deepseek r1 for high precision search engine capabilities.

**RK Industries Limited | Full stack web developer  *July 2024 - Nov 2024***

* Deployed a full-stack AWS powered enterprise scale WebRTC, RTM platform handling user authentication, dynamic group management, virtual lobby, and ultra-low latency on demand audio/video streaming.
* Composed elastic cloud infrastructure with auto-scaling and load balancing achieving 99.9% uptime through distributed architecture and redundant failover mechanisms.
* Integrated comprehensive monitoring and logging using AWS CloudWatch and CloudTrail for real time performance metrics and system health monitoring with DDoS preventive guardrails.
* Led development of multiple non-responsive company websites including the company homepage and troubleshooting network issues, VPC and domain maintenance.

**RESEARCH AND PROJECT WORK**

**Dual distilled LLM for advanced reasoning *AIMO Prize-2, 2025***

* Enhanced reasoning power of deepseek r1 distill 14b model with reinforcement learning via group relative policy optimization by using DoRA fine tuning methods and custom reward functions.
* Trained extensively on a dataset of 1.2 million math problems collected from multiple sources. Generated unique instruction prompts for all examples using Gemini 1.5 Flash for instructional fine tuning.
* Achieved tool integrated reasoning abilities to produce python code for problem solving by knowledge distillation on a dataset of 3000+ examples created by responses from 32B Deepseek model.
* Recorded scores of 25 on AIMO prize 2 leaderboard and 74.6 on AIME 2024 benchmarks, outperforming Qwen 2.5 and deepseek 32B on Live AOPS ranking metrics.

**Multilingual SLM Adaptation**

* Successfully fine-tuned Google’s Gemma models (2B and 9B parameters) for Spanish and Hindi languages using LoRA, PEFT, AWQ to achieve competitive multi-lingual capabilities against larger multilingual models of 70-75B params.
* Optimized tokenization pipelines for Hindi-specific character sets, Devanagari scripts and Spanish linguistic patterns to optimize context window utilization and reducing token fragmentation by 47%.
* Synthesized specialized datasets with prompt - chat templates comprising 100,00+ samples across diverse text corpus.

**Enhanced Neural Turing Machine *\*\* Ongoing project***

* Successfully architected a novel Neural Turing machine with 3D and 4D memory tensors with cross dimensional attention for enhanced reading and writing operations tailored for abstract reasoning tasks.
* Demonstrated competitive accuracy on ARC AGI benchmark tests compared to traditional deep learning architectures without using accelerated hardware.

**Temporal Bio Recurrence Prediction *MICCAI 2024 (LEOPARD GC)***

* Developed deep learning models predicting prostate cancer bio recurrence time frame of patients using 800 Gb of WSI (whole slide image) with data embedded in 5 levels of slides stacked in TIF, TIFF format.
* Devised custom patch-filtering algorithms to map data patches with labels of varying shapes, ensuring memory efficient data loading on TPUs to speed up parallel slide processing by 70-80% during training.
* Designed feature extraction pipeline for quantifying cellular characteristics, identifying high density cancer cell clusters from WSI data during post processing for diagnostic insights.

**Binary Encoded Multi Label 3D Segmentation *MICCAI 2024 (PENGWIN GC)***

* Developed 2.5, 2D segmentation methods with voting mechanisms to process 3D volumes efficiently, achieving exception metrics (99.79% accuracy, 98.47 IOU, 1-5 HD95) during inference and reduced computational overhead.
* Implemented ensemble of 5x U-Nets for binary encoded X-ray segmentation, processing 50,000+ images with superior performance (97.31% IOU, 3.5 HD95) on validation datasets with limited training.

**Exoplanetary Atmospheric Spectral Analysis *Ariel challenge - NeurIPS 2024***

* Created hybrid ML pipeline achieving 100ppm prediction accuracy for exoplanetary chemical spectrum analysis of AIRS and FGS sensor data from 670 exoplanets using mathematical modeling and Bayesian data analysis.
* Incorporated feature extraction pipeline combining methods of higher-order derivatives, temporal binning, linear interpolation for ground truth spectrum vectorization for efficient data cleaning.
* Assembled a dual stream architecture integrating CNN-LSTM networks for spatio-temporal analysis and custom residual networks for precise error correction.

**Multi-Modal Product Attribute Predictor *Amazon ML hackathon 2024***

* Engineered a hybrid natural language processing – computer vision pipeline to extract image features along with image captioning across 150,000+ images with lightweight BLIP, OCR models.
* Leveraged vision transformers, Llama 3.2 to perform contextual reasoning using Groq API to reduce latency, achieving accuracy of 73% of 500,000+ images on a single inference pass.

**ACHIEVEMENTS AND CERTIFICATIONS**

* **AWS (Amazon Web Services)** Certified Cloud Practitioner
* **AWS (Amazon Web Services)** Certified Solutions Architect
* 400th position **Amazon ML challenge 2024** (Top 2% out of 18,000+ teams)
* 93.69 percentile **JEE-MAINS 2022** (Top 6% out of 9 lakh candidates)
* Appointed as training and placement coordinator of Metallurgical and Materials Engineering department, Batch of 2026

**RELEVANT COURSEWORK**

* **Machine Learning Specialization** (Coursera)

Supervised Learning | Advanced Learning Algorithms | Unsupervised Learning | Reinforcement Learning | Recommenders

* **Deep Learning Specialization** (Coursera)

Convolutional Neural Networks | Sequence Models | Neural Networks and Deep learning