

Shahul Abdul

LLMs - ML - DL - Agents at Amazon | 7+ years of expertise in Machine Learning & Data Science
Fine-tuning LLMs| MCP | NLP | Gen AI | Multi-Agent Systems | RAG | | AWS | Full Stack | DSA

Building scalable e-commerce products for Customer Persona, Recommendations with Classical **ML**, **DeepLearning** and **Agentic AI**, leveraging **LLMs**, **multi-model** orchestration, context-aware workflows, **RAG** pipelines, **vector DBs**, and **statistical** analytics.

✉ abdul.shahul@outlook.com

☎ +91- 9515988770

📍 Hyderabad | Bangalore

🌐 linkedin.com/in/shahul

AMAZON

Sr. Data Scientist

Amazon

09/2022 - Present

Bengaluru, India

Roles/Responsibilities

- Spearheaded the fine-tuning and deployment of LLMs (**Deepseek, Mistral, Qwen, Falcon, BERT variants**) to enhance product attribute extraction and compatibility mapping, driving a ~15% lift in classification precision across retail catalog pipelines.
- Architected a real-time **Language Model**-based retrieval system for **Recommender Systems**, leading to a measurable increase in attach rate and *add-to-cart* conversion.
- Led GenAI model optimization strategy using **LoRA**, **quantization**, and **distillation**, enabling deployment of 10B+ parameter models under tight latency and cost constraints (~4x *throughput gain*).
- Defined and implemented scalable MLOps infrastructure using AWS **SageMaker**, **ECS**, and Lambda, powering both synchronous CX features and asynchronous enrichment workflows.
- Evaluated and integrated **vector** and **graph** databases to it power hybrid retrieval systems combining semantic relevance and compatibility logic.
- Collaboration with **Product**, **CX**, and **Engineering** teams to define LLM-in-the-loop design patterns, ensuring safe, context-aware outputs in production.
- Mentoring junior scientists and engineers in Language Model training techniques, prompt optimization, and scalable inference strategies.

TATA CONSULTANCY SERVICES

Data Scientist

Google Business Unit

02/2019 - 08/2022

Chennai, India

Roles/Responsibilities

- Built ML pipeline to classify IT service tickets as automatable, partially automatable, or manual, increasing automation coverage by ~30%.
- Developed a high-priority alert prediction model, reducing incident response delays by ~25% with **Probability & Statistics**.
- Used Naive Bayes and rule-based NLP techniques (**NLTK**, **spaCy**) for robust ticket categorization.
- Implemented multi-label classifiers to predict support routing parameters (Category, Type, Assignment Group) with > 95% accuracy.
- Managed scalable storage and processing using **BigQuery**, and deployed services via **Google Cloud Run** (CI/CD).
- Delivered real-time visual dashboards and query-driven reports, enhancing stakeholder visibility into operations.

SKILLS

Machine Learning | Deep Learning

Generative AI

DSA

LLMs

LangChain | LanGraph

python

Agents

pytorch | RAY

Transformers

MCP

Strands | Multi-Step Reasoning

Docker

OOPs

Model Context Protocol

Sagemaker | Bedrock

Encoder | Decoder

AWS | GCP

MLFlow

fastAPI

Hypothesis testing

Flask | Mongo DB

version control | git

Linux

Shell Scripting

Redis

Vannila JavaScript

AJAX | jQuery

selenium

Celery

ML System Design

E2E Project Ownership

Product-Oriented ML

A/B Testing

Pilots

Customer Experience (CX)

KEY ACHIEVEMENTS & INNOVATIONS

Prompt Optimization Agent

- Designed a prompt evaluation and tuning system (inspired by RLHF concepts) that **improved LLM answer consistency by 25%** in product attribute extraction tasks.

Built an automated auditing system

- a **multi-LLM ensemble**, leveraging proprietary language models to evaluate catalog data, reducing manual QA workload by ~**60%** and scaling audit coverage across millions of listings.

EDUCATION

Masters in Machine Learning - MCA (2022)

Sastra University - Tamilnadu

Bachelor of Science (2018)

Computer Science Specialisation degree with Mathematics and Electronics - Krishna University - Andhra Pradesh

WELL KNOWN FOR

Teaching Skills

Fail fast

Bias for Action

Tech Talks

Talent Development

Ownership