Jules Mohammed

julesmhad@gmail.com | Website | (901) 648-2857 | LinkedIn | Redmond WA

Excited to "build, ship and measure" in the Product space, software engineering, or machine learning/AI projects into production at your team. 5+ years in engineering with applicable CI/CD, Cloud services, IaaC, DevOps experience.

RELEVANT EXPERIENCE

Freelance, Software Engineer, Machine Learning, GenAI   2023 – Date
- Built a Milvus + BERT based semantic search feature into existing eCommerce platform which projected a 40% improvement in search relevance and a 30% increase in user engagement. Initial flight tests show a 2023 Q4 growth of 15% in active users.
- Built 10+ context aware customer service AI apps for clients resulting in 15%+ more customer engagement and 18%+ resolution/reduced wait times compared to prior baselines with LangChain, LangServe, OpenAI API, Streamlit (for monitoring).
- Built UI and front-end for apps interfacing chat APIs for clients using **React, JSX, Langchain.js** yielding more intuitive virtual assistants and integrated finetuned LLM service into existing VA infrastructure adding at least 25% engagement as reported.

Microsoft, Bing Ads & Serving, Bellevue WA,   [MLOps , Online Experimentation, Data Infra]
Applied Scientist II, Embedding based Ads retrieval at Audience Intelligence Platform Team,   2021 - 2023
- Built & maintain distributed data pipelines **serving 10B** users, **1B** Msft users & **0.5B** events, generating embeddings for CVR/CTR tasks using user behaviors, Ads features + Ad embedding serving, User+Ads ranking service with **Java – Spark, Kafka, SQL, Cosmos, C/C++** on **Azure.**
- **Led Income Targeting development.** Trained BERT based models which analyze petabyte scale user behaviors to predict user conversion likelihood based on income. Deployed end to end production pipelines for training, inference and generating user embeddings for Income Targeting Analytics (for campaigns) and measured several flight benchmarks/KPIs projecting revenue of $**X million**.
- Optimized existing pipelines with **CI/CD**, reducing token costs by 60% & runtime by 0.5x by efficient data sampling in ELT scripts. Refactored modules with **Airflow/Aether, ScopeSQL, C++, Cosmos**, running on Azure (**ADF, ADL, AzureML, AzureDevOps, AKS: Azure Kubernetes**)
- **Led development** for vNext frameworks for running parity analysis and quality checks to ensure 100% accuracy in GNN based embeddings in downstream tasks. Built monitoring and alerting systems resulting in 99% uptime, still in service to date.
- Successfully mentored an intern on a adjacent team to onboard quickly, who joined Microsoft Full Time in 2023.

IBM, IBM Z, Poughkeepsie, NY   [Backend, CI/CD, Performance & Databases]
Software Developer, Hyper Protect Data Controller at Z/OS Performance Team   2020 - 2021
- Built distributed data pipelines for running **performance tests** on IBM HPDC vNext resulting in 30%+ faster deploy/integration with **RDBMS e.g., PostGreSQL, OracleDB, IBM Db2** leveraging **Java, Ansible, Docker, Kubernetes, Logstash, IBM MQ.**
- Delivered NRT data pipelines providing data from regressive performance workloads run on IBM cloud which is leveraged by GTM strategy business & marketing partners. Designed **Python/Grafana/InfluxDB**, & BI dashboards to visualize timeseries performance datasets.
- Collaborate with development teams to optimize HPDC for data platforms (DMPs) improving performance and resource efficiency by 13%

BEDC Electric Plc, Benin, NG   [Backend, ML, Data Infrastructure, CDP/DMP]
Software Engineer, Data Infrastructure at Data Management Platform Team   2016 - 2020
- Built and maintain end to end (NRT online & offline) ELT pipelines serving **million user**s on the customer data platform and manage DMP running analytics for Energy Billing accounting for **$X million** gross in monthly revenue.
- **Led 3** cross-functional teams' collaboration (20+ people, 3 stakeholders) including product, platform and engineering to develop and deploy two Personalized offer pilots of BEDC Brand for maximum demand customers using **Python, AWS S3, SNS/SQS, CloudWatch, Kubernetes** Prototyped a user ranking algorithm that predicts the top likely customers to exceed prepaid usage quotas & recommend top-up by alerts.
- **Shipped the WebApp** and **Django based REST APIs** providing insights used by client technical officers for tracking deltas in customer energy usage data while resolving complaints and running B2C triages using Python**, MySQL, Chart.js, BootStrap, JavaScript.**
- **Shipped chatbots** using TDIF on FB Messenger with Flask to respond customer complaints for millions in monthly user customer base and managed data ingested into a **NoSQL** database using **MongoDB** Atlas. Support our migration to Couchbase clusters.
- **Analyzed large** multi-dimensional time series data, identifying usage patterns and key features to build predictive models which infer anomalies in energy consumption of high reactance customers saving over **$500k** in avoidable costs.
- Provides 24/7 support as well as the standard litany of DevOps tasks during on-call rotations.
- Successfully mentored interns and onboarded new hires to understand project scope quickly and avoid blockers by working scalably.

EDUCATION & RESEARCH EXPERIENCE

University of Memphis, Master of Science,   Major: Computer Engineering,   Minor: Electrical Engineering
Courses: Artificial Intelligence, Information Retrieval, Computer Vision, Image Processing, Data Mining, Deep Reinforcement Learning, NLP, NLU
Graduate Research Assistant
- Proposed PySIM: a U-Net based NN model for reconstructing 3D images from 2D layers captured from Structured Illuminated Microscopes
- Built TunableSIM GUI with C++ with Matlab's Engine API for **C/C++** and tested new features. Presented at OSI/COSI/SPIE Conference 2021
- 1st place in 2021 and 2020, at the University Research forum, two years in a row and regularly attended ML conferences.

University of Ilorin, Bachelor of Science   Major: Electrical Engineering,   Minor: Electronics Engineering
Courses: Algorithms & Data Structures, OOP, Web Mining & Search Engines, Machine Learning, Database Systems, Adv. Statistics, Optimization

POC + Side Projects

- **Bing:** Experiment to improve pCTR by including user events+ time embeddings into sequential models, to optimize ads experience.
- **CIRL:** Developed model for real time text OCR of package labels datasets for Fedex's IIS in UofM with Tesseract with accuracy > 95%.
- Deployed 7B, 13B LLM stack for 5+ client's local & offline inference and domain specific tasks, enabling efficient task automation by 30% using Ollama, Docker, Supabase pgvector, Langchain.js, JavaScript and Nextjs
- Experiment and tokenize task-specific datasets for chatbots, data search, semantic search ML apps using vectorDB**s** like Pinecone, pgVector Milvus, Pinecone, Qdrants, Weaviate for clients leveraging open data.