

Jules Mohammed

julesam47@gmail.com | [Website](#) | (470) 929-2012 | [LinkedIn](#) | Redmond WA

5+ years engineering expertise spanning Infra, Ads Retrieval using NLP techniques + GNNs, Deep Learning, Large Scale Recommender. DevOps done by collaborating cross functionally. Proficient in Python, Java, Scala, Pytorch, Tensorflow/TFX, Spark, Graph Models, Azure & AWS

RELEVANT EXPERIENCE

Stealth Startup (GenAI), **Software Engineer**, ML

May 2023 – Today

- Built a semantic search feature into an eCommerce platform which projected a **40%** improvement in search relevance and a **14%** increase in user engagement. Initial flight tests show a **2023 Q4 growth of 9%** in active users.
- Built a RAG system with a transformer-based question-answering model for a customer service chatbot to improve prior response accuracy which led to **30%** reduction in response time and a **20% increase** in customer satisfaction scores. Worked closely with 2 Senior Developers, 1 Lead PM and 2 stakeholders in a stealth startup to ensure seamless integration and deployment.

Microsoft, Bing Ads & Serving, Bellevue WA,

[MLOps, Online Experimentation, Data Infra]

Applied Scientist II, [Embedding based Ads retrieval](#) at Audience Intelligence Platform Team,

Aug 2021 – Mar 2023

- Built and maintain distributed data pipelines and GNN models **serving 5B** users, **1B** Msft users & **0.5B** events, generating embeddings for CVR/CTR tasks using user behaviors, Ads features + Ads serving, User+Ads ranking service with **Java, Spark, Kafka, SQL, Cosmos, Torch**
- **Led Income Targeting development**. Trained BERT models that analyze petabyte scale user behaviors to predict user conversion likelihood based on income. Used **LLM pipeline** for data enrichment & collaborated with Product team & stakeholders to finalize vNext, roadmap, KPIs
- 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 revenues of **\$XX million**.
- Optimized existing pipelines with CI/CD methods 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, Kubernetes**)
- Collaborated with Product team & Leadership to finalize Income Audience product vNext's roadmap, KPIs, incremental feature plans
- **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

Oct 2020 – Sep 2021

- Built automated pipelines for running **performance tests** on IBM HPDC vNext resulting in 30%+ faster deploy/integration with DBs e.g., **Postgres, MySQL, MariaDB, OracleDB, IBM Db2** leveraging tech stack: **Java, Ansible, Jenkins, Docker, Kubernetes, Terraform**
- 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 Platform Team

Nov 2016 – June 2020

- Built and maintain end to end (NRT online & offline) ELT pipelines serving **million customers** and manage data pipelines 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 access to customer data insights used by client technical officers for tracking deltas in customer energy usage 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.
- 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

- 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.
- Prototyped a document retriever and question-answering model to enhance product recommendation for a new e-commerce platform testing chat-based storefront using foundational LLMs for NLU and BERT + Milvus for efficient retrieval of product information.
- Deployed 7B, 13B LLM stack for 3+ client's local inference and domain specific tasks, enabling efficient automation by 30% using Ollama, **Docker, Supabase pgvector, Langchain.js, JavaScript and Nextjs**