Ajin Frank Justin

857-356-5917 | ajinfrankj@gmail.com | linkedin.com/in/ajin-frank-j | github/justin-aj

Education

Northeastern University **Sep 2024 - Dec 2026**
Master of Science in Data Science **GPA: 4.0/4.0**
- Coursework: Machine Learning, Web Development, Scalable Distributed Systems, MLOps, NLP, Algorithms

REVA University **Jun 2019 - Jul 2023**
Bachelor of Technology, Computer Engineering **GPA: 9.11/10**

Skills

Languages: Python (Advanced - 5yrs), TypeScript, JavaScript, GoLang, SQL (Advanced - 3yrs), C/C++, Java
Web/Backend: Node.js, Express.js, Next.js, React, Flask, FastAPI, Django, RESTful APIs, WebSocket, JWT, OAuth2
Databases: PostgreSQL, MongoDB, MySQL, Redis, Databricks, BigQuery, Pinecone
Distributed/Cloud: Apache Spark, Hadoop MapReduce, Kafka, AWS, Azure, GCP, RabbitMQ
DevOps/Tools: Docker, Kubernetes, Terraform, GitHub Actions, CI/CD, Airflow, MLFlow, Grafana, nginx, PM2

Experience

Full-Stack AI Software Engineer (RA) **Jan 2026 – Present**
DASH AI Hub, Northeastern University *Boston, USA*
- Architected scalable microservices achieving **300-600x latency improvement** (30s to 100ms) using **RabbitMQ**, **Redis**, and async worker pools, handling 100+ concurrent users with fault-tolerant queue-based design.
- Engineered RAG system with **LlamaIndex**, **Qdrant** vector DB, and **LangGraph** multi-agent orchestration; deployed via **Docker** containers with **PM2** process management for high availability across 10 instances.
- Built hybrid **Python**/TypeScript backend leveraging **PyTorch**, Transformers, and **CUDA** acceleration for ML inference, achieving 81% reduction in inappropriate outputs through AI-powered quality validation.

Machine Learning Co-op **Jun 2025 – Dec 2025**
AARP, National Office [Manager Recommendations] *Washington DC, USA*
- Developed gradient boosting model processing **5M+ samples** in **PySpark MLLib**, achieving +18% conversion lift and +12% retention uplift through personalized digital offer scoring.
- Built ML monitoring framework in **Databricks** tracking AUC, KS, lift across **25+ production models**; integrated statistical drift detection reducing model degradation incidents by 40%.
- Engineered scalable feature pipeline processing **3,000+ features** with model-based selection using **Spark** distributed computing, improving F1 and AUC-PR by 5-8%.

Software Engineer **Jun 2023 – Jun 2024**
Dynapac, Fayat Group [Manager Recommendations] *Bangalore, India*
- Designed ETL pipeline processing **300M+ telemetry records** from 1000+ nodes using **Python** batch processing with **Azure** Durable Functions, transforming raw data to structured JSON in Blob Storage.
- Led database optimization transforming multi-join PostgreSQL architecture into partitioned schema, achieving **200% read performance improvement** validated with pgBench benchmarking.
- Conducted time series forecasting using **ARIMA** models on operational data, improving ROI by 20%; automated CI/CD pipelines with **Docker** containerization for zero-downtime deployments.

Projects

GoodReads Clone – Full-Stack Book Platform | *Next.js, TypeScript, Node.js, MongoDB* [link] **Sep 2025**
- Built full-stack app with **Next.js**/**TypeScript** frontend, **Node.js**/**Express.js** RESTful API, **MongoDB** persistence; implemented authentication, Google Books API integration, deployed on **Vercel**.

AskNEU RAG System | *LangChain, Docker, Kubernetes, Pinecone, GCP* [link] **Jan 2025 – Apr 2025**
- Architected RAG system with Cohere reranking, **LangGraph** orchestration; scraped **50,000+** pages into **Pinecone**; deployed with **Docker**, **Kubernetes**/**Terraform**, **Airflow** DAGs, and **Grafana** monitoring.

Kambaz Learning Management System | *Next.js, TypeScript, Node.js, Express* [link] **Sep 2025 – Dec 2025**
- Developed Canvas LMS clone with **Next.js** SPA, **Node.js**/**Express.js** backend; implemented RBAC for 3 user roles, JWT authentication, and scalable CRUD APIs for courses, assignments, and grading.

Food Categorization ML System | *Python, scikit-learn, TF-IDF, PCA* [link] **Oct 2024 – Dec 2024**
- Developed **scikit-learn** ML pipeline classifying **1.7M USDA entries** into 70+ categories achieving **91.98% accuracy**; optimized with TF-IDF vectorization, PCA dimensionality reduction, and A/B testing.