BHARATHRAAJ NAGARAJAN

Los Angeles, CA | +1 716-228-7052 | bharathraajnagarajan@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

Machine Learning Engineer with 3+ years of experience designing and deploying ML and NLP systems at scale, including transformer-based sequence-to-sequence models and end-to-end training pipelines. Strong hands-on experience with PyTorch, TensorFlow, Spark, and production operations on Kubernetes with 99.9% uptime. Proven impact improving model accuracy, reducing release cycles, and operationalizing ML solutions across enterprise environments on AWS, Azure, and GCP.

EDUCATION

University at Buffalo, The State University of New York **Aug 2024 - Dec 2025**
Master of Professional Studies, Data sciences and Applications (GPA: 3.9/4.0)
- **Coursework:** Database Management Structure, Data Intensive Computing, Machine Learning, Data Modeling, Cybersecurity

Anna University - Sri Sivasubramaniya Nadar College of Engineering **Aug 2017 - May 2021**
Bachelor of Engineering, Minor in Computer Science

WORK EXPERIENCE

Tata Consultancy Services | *SYSTEMS ENGINEER (Machine Learning / Big Data AI)* **Oct 2021 - Apr 2024**
- Developed and deployed transformer-based NLP and sequence-to-sequence models in PyTorch, improving predictive accuracy by 18% across enterprise-scale use cases.
- Designed and built end-to-end ML pipelines in TensorFlow, covering data ingestion, feature engineering, model training, evaluation, and deployment, reducing release cycles by 28%
- Deployed and monitored ML services on Kubernetes, maintaining 99.9% production uptime while supporting scalable model serving and experimentation.
- Built scalable data/feature pipelines leveraging Kafka, HDFS, HBase, and Kudu, reducing query latency by 40% in high-throughput environments supporting real-time analytics and ML workloads.
- Optimized PySpark and Spark SQL pipelines supporting model data preparation and batch workflows, reducing ETL processing time by 47% through performance tuning and execution optimization.
- Collaborated with cross-functional stakeholders to translate business requirements into production ML solutions using BigQuery, Scikit-learn, Azure, and Airflow, building end-to-end pipelines that accelerated deployment and improved model reliability

U.S. INTERNSHIP EXPERIENCE

Media Sales Plus Inc. | *Data Engineering Intern, Analytics | Buffalo, NY* **Aug 2025 - Dec 2025**
- Developed an executive-facing Streamlit analytics platform with KPI monitoring, anomaly detection, and ARIMA-based forecasting to support recurring financial reporting.
- Built Python ETL pipelines to ingest, validate, aggregate, and backfill client datasets; standardized schemas to improve consistency for analytics and ML-driven insights.
- Implemented CI/CD using GitHub Actions and integrated workflows with Azure cloud services to enable secure, repeatable deployments and automated execution.
- Implemented a rule-based NLP chatbot interface that translated natural-language business questions into computed metrics within the analytics application.

Selected Engineering Project

Production-Grade ML/Streaming Systems Reliability (PySpark + Structured Streaming) | Link **Jun 2025 - Aug 2025**
University at Buffalo, The State University of New York
- Built a PySpark shortest-path optimizer using Dijkstra's algorithm over large graphs (RDD/DataFrame) with CLI-based validation.
- Implemented a fault-tolerant Spark Structured Streaming pipeline with checkpointing (state + offsets) and micro-batch recovery.
- Debugged streaming reliability issues (checkpoint concurrency, sink/output-mode compatibility) to deliver a reproducible end-to-end workflow.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, SparkSQL, HiveQL, R, Bash, Scala
- **ML/AI:** PyTorch, TensorFlow, Keras, Scikit-learn, Transformers, NLP, CNN, OpenCV, Reinforcement Learning (exposure), Prompt Engineering
- **Data Engineering:** Hadoop, Apache Spark, Apache Flink, Apache Kafka, HDFS, Hive, HBase, Kudu
- **BI/Analytics:** Power BI, Tableau, Looker, Excel, Matplotlib, Plotly
- **Cloud & Data Platforms:** AWS, Azure, GCP, Microsoft Fabric, Databricks, Snowflake, BigQuery, PySpark
- **MLOps / Deployment:** Kubernetes, Docker, CI/CD (GitHub Actions)
- **Tools:** SSIS, SSAS, SSRS, Power Automate, Airflow, dbt, Alteryx, GIT, Docker, GitHub Actions
- **Certifications:** AWS Certified Solutions Architect - Associate (In Progress)