

ALI HAIDER

Senior AI/ML Engineer | LLM & Agentic Systems | Production ML at Scale

Kansas City, KS

📞 984-218-8048 ✉️ malih4110@gmail.com in ali-haider

🌐 muhammadali4.github.io 🐙 github.com/MuhammadAli4



SUMMARY

Senior AI/ML Engineer with 6+ years building and deploying production machine learning and generative AI systems across finance, retail, telecom, and cybersecurity. Deep hands-on experience designing multi-agent LLM architectures (LangGraph, LangChain), fine-tuning transformer models, and operating conversational AI at scale (1M+ calls/day). Proven track record shipping end-to-end ML pipelines on AWS/Docker/Kubernetes, leading engineering teams, and translating model performance into measurable business outcomes for stakeholders from engineering to executive leadership.

EXPERIENCE

AI/ML Engineer

Kansas City, KS

CAPGEMINI | CLIENT: CHARTER SPECTRUM

May 2026 – Present

- Deploy and optimize LLM-driven conversational AI agents on the Kore.ai platform in production at scale, supporting roughly 1M IVR calls per day for Charter Spectrum's customer service operations.
- Leading the technical migration from Kore.ai to an in-house conversational AI platform, architecting the NLU, dialog management, and orchestration layers that will replace the third-party solution, eliminating Kore.ai platform/licensing costs for a projected \$4M in annual savings.
- Partner with Google on an NGA initiative to bring natural-sounding, human-like generative voice to IVR interactions, working across model selection, latency, and voice quality tradeoffs.

Founding AI/ML Engineer

Remote

APPLY SPRINT - (ACQUIRED)

Aug 2025 – May 2026

- Architected and shipped a production multi-agent AI system (LangGraph + LangChain) for automated job matching and resume optimization, orchestrating GPT-4, Claude, and DeepSeek across a structured agent pipeline.
- Designed a two-stage NLP pipeline: fine-tuned XLM-RoBERTa (PyTorch) as an upstream classifier feeding a downstream stacked ensemble, reaching 90% accuracy predicting LinkedIn Easy Apply eligibility from raw job descriptions, a core product differentiator.
- Deployed the full system end-to-end on AWS via Dockerized containers, and recruited, led, and managed a small engineering team across architecture, ML pipeline design, and production deployment through to company acquisition.

Senior Data Scientist

Charlotte, North Carolina

BANK OF AMERICA

Feb 2024 - Aug 2025

- Led end-to-end model validation and ongoing monitoring for large-scale classification models supporting Mastercard and OMNI commercial credit card portfolios, covering \$20B+ in annual transactions. Monitored model performance using CSI, PSI, AUC-ROC, and rank ordering, ensuring stability and regulatory compliance.
- Conducted deep statistical root cause analysis across dozens of KPIs, and developed surrogate fraud detection models using black-box distillation techniques to gain insights into the third party models.
- Authored executive-facing model documentation and risk reports communicating performance findings and remediation strategies to senior leadership and cross-functional stakeholders.

Data Scientist

Charlotte, North Carolina

LOWE'S

Feb 2022 – Jan 2024

- Media Mix Modeling: Designed and implemented a robust statistical framework to evaluate marketing effectiveness. Built a baseline sales model using XGBoost to isolate incremental impact, tackled multicollinearity, and applied ensemble and Bayesian (PyMC3) methods for accurate attribution. Conducted extensive data analysis and experiments to validate assumptions, handle lag effects, and overcome data sparsity. Resulted in a 3% uplift in ROI on a \$900M annual marketing spend.
- Sales Prediction During Extreme Weather: Trained, tested, and deployed regression models to forecast sales across 3,000 stores and 30,000 assortments during severe weather disruptions. Used K-Means clustering to group similar stores and product categories, enabling tailored modeling strategies per cluster. Integrated weather data, historical sales, and store-level metadata to enhance feature engineering. Leveraged scikit-learn, XGBoost, and hyperparameter tuning for model optimization. Deployed full end-to-end ML pipeline using Docker, Kubernetes, and MLflow for

experiment tracking and model registry, with CI/CD managed via GitHub. Improved inventory planning, reduced stockouts, and boosted customer satisfaction through data-driven decisions.

Data Scientist

Irvine, California

HTN NETWORKS INC CISCO (CONTRACT)

Apr 2020 – Jan 2022

- Developed and implemented a reinforcement learning and ensemble machine learning model to identify malicious network traffic. Continuously improved model performance through experimentation and testing. Established robust post-deployment checks, including data drift monitoring, to guarantee ongoing effectiveness.
- Improved model accuracy through rigorous testing, leading to precise threat identification. Implemented post-deployment checks ensuring model reliability and ongoing effectiveness, achieving a 97% F1 score, reducing false positives by 42%, and improving response time by 35%.

Graduate Research Assistant

Raleigh, North Carolina

NC STATE UNIVERSITY

Aug 2017 – Jan 2018

- Contributed to a sustainability research lab, applied ANOVA and ARIMA for trend forecasting, enhancing prediction reliability to analyze environmental data trends for environmental impact assessments.
- Utilized statistical methods to derive insights, improving environmental impact assessments.

Teaching Assistant

Istanbul, Turkey

ISTANBUL SEHIR UNIVERSITY

Sep 2015 – May 2017

- Supported undergraduate instruction through grading, tutoring, and leading discussion sessions.
- Led Python and SQL lab sessions, improving student coding proficiency and understanding for assignments related to industrial engineering and Machine learning.

EDUCATION

North Carolina State University

Raleigh, NC

MASTER OF SCIENCE IN OPERATION RESEARCH

Dec 2020

Istanbul Sehir University

Istanbul, Turkey

MASTER OF SCIENCE IN INDUSTRIAL & SYSTEM ENGINEERING

Dec 2017

NED University of Eng and Tech.

Karachi, Pakistan

BACHELOR OF ENGINEERING IN INDUSTRIAL ENGINEERING

Dec 2013

KEY SKILLS

LangGraph, LangChain, Multi-Agent Orchestration, GPT-4, Claude, DeepSeek, RAG, Prompt Engineering, Transformers (XLM-RoBERTa), Hugging Face, PyTorch, TensorFlow, XGBoost, Bayesian Modeling (PyMC3, PyStan) IVR/Dialog Systems, NLU, Generative Voice, AWS, Docker, Kubernetes, MLflow, CI/CD, Airflow, Git, Python (Pandas, NumPy, SciPy), Advanced SQL, Snowflake, Time-Series Forecasting, A/B Testing, Statistical Analysis, Technical Team Leadership, Executive Communication, Cross-Functional Stakeholder Management

RECENT APPLIED PROJECTS

- Autonomous AI content system, built a three-engine (Strategist / Publisher / Analyst) architecture that generates, publishes, and self-optimizes an AI-avatar YouTube Shorts channel, including a video lip-sync pipeline built on LTX-2 generative video models and a YouTube Analytics feedback loop driving content strategy.
- Quantitative trading research, designed and backtested 1,000+ rule-based strategy variants on QQQ/GLD under strict train/holdout methodology to evaluate risk-adjusted returns versus buy-and-hold benchmarks.
- SEO & growth automation, built an automated keyword discovery and programmatic landing-page pipeline plus a targeted outreach scraper, driving organic top-of-funnel growth for a B2C SaaS product.

PUBLICATIONS

Authored 7 peer-reviewed publications in reputable journals, leveraging statistical methods and data science techniques to extract insights and solve real-world problems. Full list available at: **Google Scholar:** <https://scholar.google.com/citations?user=tq8Rh54AAAAJ&hl=en>

HONORS & AWARDS

2017 **Recognized with NC State PhD fellowship award**, A prestigious award granted to a select few pursuing a PhD, recognizing exceptional academic excellence and research potential. NC, USA