Michael Long

www.linkedin.com/in/michael-long88/ | www.mlong.me

Marine Corps veteran and technical leader with 5+ years building production systems, including 2+ years specializing in NLP/LLM applications. Transitioned from software engineering to data science, bringing practical expertise in full-stack development and production deployment. Led development of RAG systems, automated document processing, and LLM evaluation frameworks for research and government clients. Proven track record of significant performance improvements through engineering-driven approach to machine learning.

TECHNICAL SKILLS

Programming Languages: Python, SQL, Rust, JavaScript, R
Machine Learning: Large Language Models (LLMs), Retrieval Augmented Generation (RAG) Systems, Natural Language Processing (NLP), Cross-encoders, Vector Databases, Clustering, Regression, Random Forest, XGBoost
Machine Learning Frameworks/Tools: HuggingFace Transformers, Ollama, LanceDB, spaCy
Back End Development: FastAPI, Node.js, CKAN
Front End Development: Vue, Bootstrap, Tailwind
Cloud-Based Applications: Amazon Web Services (AWS)
Dev-Ops and Containerizations: Linux, Docker, Git

EXPERIENCE

RTI International   Raleigh, NC
Research Data Scientist   May 2023—Present

- Led development of LLM-powered transcript processing system with Vue frontend, FastAPI backend, and vector database integration, reducing manual coding time from 4 hours to 30 minutes per transcript through automated PDF extraction and course standardization.
- Used cross-encoder models and SANDS framework for automated supplier name resolution across energy billing data from several thousand suppliers nationwide, achieving 86% accuracy for EIA's national energy consumption analysis.
- Created style and efficiency guides for CDC microsimulation projects, implementing containerization and optimization processes that resulted in 10x speed increase for simulation execution across multiple health modeling initiatives.
- Led team of 5 data scientists analyzing public housing smoke-free policy impact in Florida, applying difference-in-differences methodology with synthetic populations and logistic regression modeling.
- Developed LLM pipeline for automated MeSH term tagging of research abstracts, reducing processing time from 30 minutes to 30 seconds per abstract for National Library of Medicine classification tasks.

NC State University   Raleigh, NC
Data Scientist   September 2022—April 2023

- Conducted network and sentiment analysis on one year of social media forum text data.
- Built machine learning models using Python to predict forum thread length and duration based on spaCy-extracted entities and graph network-based user interactions.

RTI International   Raleigh, NC
Software Engineer   January 2020—June 2022

- Built multi-tenant social media surveillance platform processing daily data from Twitter, Instagram, Reddit, and YouTube APIs with JavaScript dashboards and custom Highcharts visualizations serving multiple research clients.

- Led technical architecture for criminal justice analytics platform aggregating 911 call data from 5 cities, converting legacy Go codebase to Python for improved maintainability and implementing automated data ingestion from police department APIs.
- Developed and open-sourced Vue.js plugins Harness-Vue and Harness-Vue-UI, contributing new features, documentation, and unit tests to internal data visualization library adopted across RTI projects.
- Created FastAPI backend template for internal full-stack development, establishing coding standards and architectural patterns used by development teams organization-wide.

CGI   Raleigh, NC
Software Engineer   February 2019—January 2020

- Created Python unit test framework for Kafka-based parser program.
- Refined the quality assurance process for a legacy Java application that transmits health insurance claim images.

United States Marine Corps   US
Public Affairs Liaison, Barracks Manager, Squad Leader, Team Leader   July 2006—October 2014

- Led and mentored teams ranging from 4 to 12 Marines across multiple leadership roles, developing project management and cross-functional coordination skills in high-pressure environments.
- Created operational policies and procedures for facility management and team coordination, including scheduling systems for 150+ personnel and regulatory compliance frameworks.
- Coordinated public events and stakeholder communications with external agencies and international military partners, translating complex operations into clear messaging for diverse audiences.

EDUCATION

Master of Science in Analytics   May 2023
Institute for Advanced Analytics, NC State University, Raleigh, NC

Bachelor of Science in Computer Science | Magna Cum Laude | Minor: Mathematics   December 2018
University of North Carolina at Charlotte