

Michael Long

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

Marine Corps veteran and data scientist with expertise in LLM applications, machine learning, and full-stack development. Specialized in developing RAG-based tools, NLP pipelines, and machine learning models that deliver tangible business value. Combines technical proficiency in Python, SQL, and cloud technologies with strong leadership skills developed through military service. Committed to leveraging data science to solve challenging problems and drive innovation.

TECHNICAL SKILLS

Programming Languages: Python | SQL | Rust | JavaScript | HTML/CSS | R

Machine Learning: Large Language Models (LLMs), Natural Language Processing (NLP), Clustering, Regression, Random Forest, XGBoost

Back End Development: FastAPI, Flask, Node.js, CKAN

Front End Development: Vue, Bootstrap, Tailwind

Web-Based Data Visualization: Highcharts, Amcharts

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

- Built a RAG-based question-and-answer tool for national survey data using LLMs, ensuring concrete data-backed responses with minimal hallucinations. The tool is built using a FastAPI backend, Vue frontend, LanceDB vector database, and serves a local LLM through Ollama.
- Created style and efficiency guides for advanced CDC microsimulation projects, streamlining templates for simulation and containerization, resulting in a 10x speed increase for running.
- Developed an LLM pipeline for automating the tagging of abstracts with National Library of Medicine (NLM) Medical Subject Headings (MeSH) terms, reducing the time needed per abstract to ~30 seconds.
- Led a team of 5 data scientists on a project analyzing the impact of the public housing smoke-free policy in Florida.
- Developed an education transcript scraping tool with Vue frontend, FastAPI backend, and LLM integration via Ollama to enhance automated transcript processing and course standardization.

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

- Converted Go code to Python for ETL procedures to provide better readability and maintainability.
- Developed features for, and helped open-source, the Vue.js plugins Harness-Vue and Harness-Vue-UI.
- Developed FastAPI backend template for internal full-stack project use.
- Built JavaScript dashboards to communicate data to stakeholders.

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

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	